Libro bianco per il riutilizzo dell'informazione del settore pubblico

EVPSI Extracting Value from Public Sector Information
Progetto di ricerca finanziato dalla Regione Piemonte | http://evpsi.org

I contenuti del Libro bianco per il riutilizzo dell'informazione del settore pubblico sono stati per la gran parte ispirati dai prodotti di ricerca e dalle occasioni di confronto e discussione promossi nell'ambito del progetto **EVPSI** (acronimo di *Extracting Value from Public Sector Information*, http://www.evpsi.org), finanziato dalla Regione Piemonte, attivo tra il 2009 e il 2012. L'elenco delle istituzioni e dei ricercatori membri di EVPSI è disponibile all'indirizzo http://www.evpsi.org/content/membri.

UNIVERSITÀ DEGLI STUDI DI TORINO

dsg Dipartimento di Scienze Giuridiche

Nexa Center for Internet & Society

FONDAZIONE ROSSELLI

In collaborazione con:

AMEDEO AVOGADRO università degli studi del piemonte orientale

Progetto di ricerca finanziato da:

REGIONE PIEMONTE

Il Libro bianco per il riutilizzo dell'informazione del settore pubblico è stato redatto a cura di Federico Morando con il supporto di Raimondo Iemma e Claudio Artusio, al cui lavoro si aggiungono contributi puntuali di Mauro Alovisio, Eleonora Bassi, Juan Carlos De Martin, Alessandro Mantelero, Marco Ricolfi, Angelo Maria Rovati, Margherita Salvadori e Cristiana Sappa.

Progetto grafico: Mirko Balducci - Linked Open Data Italia.

Linked Open Data Italia
www.linkedopendata.it

Versione 1.0 (alpha) rilasciata il 24 aprile 2012
Versione 1.0 (beta) rilasciata il 12 maggio 2012

Questa opera è distribuita con licenza Creative Commons Attribuzione 3.0 Italia

Tu sei libero:

- di riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare quest'opera
- di modificare quest'opera
- di usare quest'opera per fini commerciali

Alle seguenti condizioni:

Devi attribuire la paternità dell'opera nei modi indicati dall'autore o da chi ti ha dato l'opera in licenza e in modo tale da non suggerire che essi avallino te o il modo in cui tu usi l'opera.

Indice

Voi... (target audience)

...e Noi (nota sugli autori)

Executive summary

1 Parte prima:
Il valore dell'informazione detenuta dal settore pubblico

1.1. Introduzione: la PSI come opportunità

1.1.1. Lo scenario internazionale

1.1.2. Le tendenze emergenti...

1.1.3. ...e gli errori da evitare

1.2. La PSI come piattaforma di innovazione (tecnologica e sociale)

1.3. La tecnologia

1.4. L'accessibilità dei dati

1.5. Una (prima) raccomandazione: "Dati grezzi aperti subito!"

1.5.1. Dati aperti

2 Parte seconda:
Apriamo un po' di dati

2.1 Le migliori pratiche

2.2. I passi chiave nel processo di apertura

2.3. Licenziare i dati: pillole di diritto della proprietà intellettuale

2.4. Rendere i dati disponibili: pillole di tecnologia

3 Parte terza:
Sfide, ostacoli e ulteriori opportunità

3.1. Le normativa sul riutilizzo dell'informazione del settore pubblico

3.2 Su costi, profitti e tariffe

3.3. Tutela dei dati personali: tra accesso e riutilizzo

3.2.1. L'informativa

3.2.2. L'anonimizzazione

3.3. Controlli, abusi e responsabilità

3.3.1. Obblighi o raccomandazioni? Registrazione obbligatoria degli utenti? "No, grazie."

3.3.2. Responsabilità (dai disclaimer più prudenti alla certificazione del dato)

3.4. Si può fare di più (tecnicamente)... dati aperti a cinque stelle!

E ora?

Voi... (target audience)

Questo libro è stato scritto pensando in particolare a quei politici, amministratori e funzionari pubblici, che si sentono custodi di un grande tesoro di informazioni e conoscenza, accumulato dai cittadini, con risorse dei cittadini e nell'interesse del cittadini. Ma anche a tutti gli altri dipendenti, consulenti e collaboratori delle amministrazioni e istituzioni pubbliche di ogni genere, che almeno una volta ogni tanto decidono che valga la pena, nell'interesse della collettività, cercarsi una "gatta da pelare" che nessuno gli aveva appioppato (e che, al contrario, richiede anche di rompere le scatole a colleghi e responsabili, affinché "vi facciano il favore di lasciarvi fare del lavoro extra"). Insomma, "Voi", in questo libro, siete le persone che lavorano generosamente per la "cosa pubblica".

L'obiettivo di questo libro è aiutarvi a comprendere quali opportunità la miniera rappresentata dalle informazioni pubbliche possa già offrire, ma anche quali accorgimenti adottare per aumentare simili opportunità in futuro. Al contempo, questo testo vuole evidenziare alcuni possibili rischi, prevenire alcuni errori comuni e, in generale, aumentare la consapevolezza relativa alla gestione dei dati pubblici. Questo testo non è un manuale operativo, anche se contiene alcuni spunti che speriamo possano essere utili per scriverne uno, magari arricchendo ulteriormente il manuale dell'Associazione DataGov.it [rilasciato il 14 Aprile 2011] o l'Open Data Handbook di Open Knowledge Foundation, in corso di traduzione in italiano.

Inoltre, speriamo che il nostro lavoro sia un utile supporto per quei cittadini e per quegli imprenditori che possono "dar vita" ai dati pubblici, creando opportunità di crescita – economica e in termini di partecipazione democratica. Per questi soggetti, il nostro libro potrebbe (speriamo!) rappresentare una collezione di idee e suggerimenti, utili a informare l'interlocutore pubblico curioso e a rassicurare quello dubbioso.

...e Noi (nota sugli autori)

"Noi", gli autori di questo libro, siamo un gruppo di ricercatori di istituzioni di ricerca pubbliche e private: l'Università di Torino, il Politecnico di Torino, la Fondazione Rosselli, l'Università del Piemonte Orientale. Siamo un gruppo eterogeneo, così come è richiesto dall'eterogeneità dell'oggetto della nostra ricerca: tra noi si trovano molti giuristi (dagli esperti di diritto d'autore a quelli di privacy, passando per il diritto amministrativo, il diritto della concorrenza e quello internazionale), economisti (che si occupano di economia dell'innovazione, industriale, pubblica o della cultura), ingegneri informatici (esperti di reti, di semantica, di Internet e multimedia in generale).

Molti di noi, oltre al lavoro di ricerca teorica e applicata, sono stati coinvolti in progetti concreti sui dati aperti, in particolare nell'esperienza – allora pionieristica per l'Italia – del portale www.dati.piemonte.it, lanciato nel 2010. Molti sono anche coinvolti nel network tematico europeo LAPSI (Legal Aspects of Public Sector Information, www.lapsi-project.eu).

Inoltre, più o meno direttamente, tutti noi abbiamo potuto far nostra l'esperienza e il lavoro di molti ricercatori e attivisti del mondo open data a livello nazionale e internazionale. Una sintesi particolarmente efficace della conoscenza accumulata da questa "comunità" è confluita, tradotta in termini operativi, nell'Open Data Handbook (http://opendatahandbook.org/en/) della Open Knowledge Foundation, che dunque merita qui un'esplicita menzione.

Executive summary

Siamo nella società dell'informazione. La nostra è un'economia dell'informazione. Il digitale sta cambiando tutto. Questi mantra vengono ripetuti sino alla noia, ma prendiamoli sul serio e chiediamoci cosa ciò comporti nell'Agenda (Digitale) di un amministratore o dirigente pubblico. Cosa si potrebbe fare per adeguare i meccanismi di funzionamento della macchina pubblica a questo nuovo ambiente, cosa per massimizzare le opportunità di benessere sociale per tutti? Tra le molte possibili risposte, questo libro bianco ne esplora una, suggerita anche dall'Agenda Digitale della Commissione Europea: **"mettere a disposizione le informazioni relative al settore pubblico in modo trasparente, efficace e non**

discriminatorio, [come] fonte importante di crescita potenziale di servizi online innovativi". Forse non a caso, il primo dei 24 *workshop* della prima *Digital Agenda Assembly* – gli stati generali della politica dell'innovazione in Europa – si intitolava proprio *Open data and re-use of public sector information.*

L'informazione del settore pubblico, o ***public sector information*** **(PSI)** nella dicitura anglosassone, **è varia quanto e più delle stesse pubbliche amministrazioni.** Questi dati vanno **dalle immatricolazioni delle automobili agli orari delle farmacie**, dai piani regolatori alle mappe della protezione civile, dalle rilevazioni meteo dell'Aeronautica Militare agli orari di passaggio in fermata dei mezzi pubblici.

Ovviamente, c'è una piccola parte di dati pubblici che devono rimanere riservati o addirittura segreti (per ragioni di sicurezza nazionale o di privacy, ad esempio). Non pochi dati, però, sono già teoricamente accessibili a chiunque e moltissimi altri potrebbero esserlo. **Aprire effettivamente questi dati e chiarire che essi sono liberamente riutilizzabili da chiunque è l'obiettivo dell'approccio (e del "movimento")** ***Open (Government) Data***.

L'opportunità è di portata epocale. Non solo per **favorire l'innovazione** da parte dei privati, e la **partecipazione democratica** da parte dei cittadini, ma anche per **aumentare l'efficienza della stessa pubblica amministrazione**, sia in termini di riduzione degli sprechi sia in termini di servizi per i cittadini nuovi e migliori. E i costi sono contenuti nel breve periodo e probabilmente nulli nel medio/lungo – anzi, la speranza di generare risparmi non è fuori luogo. Questo genere di prospettive fa ovviamente sorgere perplessità: **"Più servizi con meno costi: è possibile?"**. **Fortunatamente sì**, ed un esempio banale, ma efficace è fornito dagli orari dei mezzi pubblici.

Molte aziende municipalizzate realizzano siti, che mostrano ai cittadini come spostarsi coi loro mezzi. Comprensibilmente, però, non sempre questi siti sono facili da trovare e di rado l'interfaccia degli stessi è tradotta in decine di lingue diverse a beneficio dei turisti. Inoltre, i percorsi finiscono con la tratta di competenza della specifica municipalizzata. Infine, di rado questi servizi sono disponibili sotto forma di comode applicazioni per dispositivi mobili, dall'iPhone ad Android, passando per gli smart-phone di ogni marca e modello. In effetti, giustificare la spesa necessaria a realizzare questo tipo di servizi sarebbe difficile (ad esempio, un Assessore ai trasporti potrebbe chiedersi: "Dobbiamo spendere 10.000 Euro per fare un'applicazione che permetta a chi è abbastanza ricco da comprarsi un iPhone di perdere 5 minuti in meno alla fermata del tram, oppure li spediamo in buoni taxi per gli invalidi?"). Fortunatamente, le società che condividono i dati sugli orari dei propri mezzi pubblici possono star certe che queste informazioni saranno accessibili su GoogleMaps, con le sue interfacce in molte lingue (e con la possibilità di creare percorsi tra comuni e regioni diverse), e che saranno riutilizzate dagli sviluppatori di applicazioni per tutti i più comuni dispositivi mobili (alcuni dei quali offriranno anche versioni gratuite, magari per farsi conoscere o anche per invogliare all'acquisto della versione *premium* a pagamento). E **questa non è una futuribile possibilità, è quello che già comincia a succedere nelle grandi città italiane**, dove c'è così tanta fame di questi dati da utilizzare in applicazioni specifiche, che alcuni programmatori scaricano (con grande fatica e in modo più o meno lecito, anche interpretando in maniera semi manuale dei formati che non sono immediatamente fruibili) anche informazioni che nessuno si è mai preoccupato di mettere formalmente a disposizione per il riutilizzo.

Molti altri esempi (ed alcuni mini casi di studio) sono sparsi nel resto di questo libro bianco. A livello macroeconomico, gli

studi internazionali compiuti in questo campo concordano nell'affermare che il valore indotto da dati pubblici riutilizzabili, per un paese simile all'Italia, sia nell'ordine delle centinaia di milioni di Euro. Di contro, il costo di realizzare un progetto e un portale per la condivisione di questi dati può andare da qualche decina ad alcune centinaia di migliaia di Euro (con 200-300.000 Euro si può far partire un rispettabile progetto di livello nazionale; la Commissione Europea ha recentemente aperto un bando per realizzare il suo portare open data a un costo che non superi gli 800.000 Euro). In sintesi: **il gioco vale la candela**.

Come se questo non bastasse, esistono *communities online*, enti non governativi (e centri di ricerca) ben disposti ad aiutare chi voglia aprire i propri dati. E ci sono amministrazioni locali che sono già partite, anche in Italia, e che possono mettere a disposizione le loro soluzioni per il riutilizzo. Con determinazione e passione, **è addirittura possibile avviare un progetto open data a costo praticamente nullo** (in termini di specifici stanziamenti di budget e abbastanza contenuto anche in termini di ore/uomo).

Questo libro bianco, oltre ad offrire ulteriori elementi per sostanziare le affermazioni di cui sopra, cerca di fornire alcuni **principi cardine** e chiavi di lettura, fugando al contempo i più comuni fraintendimenti.

Un punto chiave, ad esempio, è che l'open data riguarda la messa a disposizione dei dati, perché *altri* possano riusarli. **Open data**, quindi, **non significa spingere la pubblica amministrazione a sviluppare nuovi servizi a partire dai dati**. Il servizio che la pubblica amministrazione deve sviluppare è "semplicemente" un modo pratico ed efficace per pubblicare i dati in proprio possesso (preceduto da un razionale processo per scegliere quali dati possano essere messi a disposizione di tutti e corredato da una chiara affermazione relativa al fatto che il riutilizzo dei dati è libero). L'esperienza mostra che i cittadini riceveranno nuovi servizi, sia gratuiti che a pagamento; ma saranno i singoli sviluppatori, gli imprenditori, le aziende e la società civile a produrre questi servizi. E il processo di apertura dei dati avrà successo se la pubblica amministrazione potrà concentrarsi a far meglio le attività essenziali che è suo compito istituzionale svolgere.

Un aspetto da non trascurare riguarda anche la narrativa relativa ai dati: **i dati non dovrebbero essere percepiti come una proprietà dell'ente, ma come un bene pubblico, un patrimonio dei cittadini.** L'aspetto dirompente delle informazioni digitali è che questo bene pubblico può davvero essere usato da tutti, senza impedire ad altri di fruirne allo stesso tempo. I dati sono come un'autostrada in cui non si formano mai code o una sorgente che non si secca mai. Per questo, è cruciale comprendere, che **i dati sono veramente aperti se e solo se possono essere liberamente riutilizzati e ridistribuiti da chiunque, per qualunque fine**. In buona fede, qualcuno potrebbe pensare che la natura di bene pubblico dei dati suggerisca di metterli a disposizione solo per scopi non-di-lucro e coerenti con l'interesse generale (ad esempio, solo per scopi educativi o socialmente utili). Questo, tuttavia, è un errore dettato dal fatto che la maggior parte dei beni pubblici materiali si consumano o si congestionano con l'uso, per cui l'uso stesso va regolato con attenzione. I dati digitali sono diversi: **il modo per garantire che essi restino un bene pubblico è renderli disponibili a tutti con le minime barriere possibili, in modo che chiunque abbia pari opportunità di riusarli**. Se qualcuno fornirà un servizio che genera anche un profitto, meglio per lui, ma anche meglio per gli altri cittadini (che – se pagano il servizio – significa lo avranno trovato utile) e per lo stato (che incamererà delle tasse su questo profitto e sull'indotto generato dal servizio). E qualunque altro soggetto potrà accedere agli stessi dati per copiare il nuovo servizio, così

creando una spinta competitiva verso ulteriori innovazioni (e/o costi più bassi per gli utenti).

Un altro principio un po' contro intuitivo è quello per cui **se un dato sembra inutile al suo detentore, è un'ottima idea condividerlo col resto del mondo**, giacché qualcun altro avrà probabilmente un'idea per utilizzarlo meglio. Per altro, spesso quell'idea potrebbe richiedere la combinazione con altri dati, anch'essi apparentemente inutili, se considerati in isolamento.

Ovviamente, perché i dati siano riusabili, una condizione imprescindibile è che questi siano facilmente accessibili, nel modo più completo possibile. In pratica, ciò significa **pubblicare i dati su Internet, possibilmente così come sono usati dalla pubblica amministrazione** stessa e lasciando che possano essere scaricati senza ostacoli che possano essere evitati (p. es. senza registrazioni). **Ma l'accesso è solo una precondizione** – con desiderabili effetti collaterali sulla trasparenza! – e non lo scopo finale del processo di apertura. **Lo scopo finale è il riutilizzo dei dati**. E siccome lo scopo deve essere il riutilizzo, ovvero l'elaborazione e la combinazione con altri dati, **i dati devono essere disponibili in un formato comodo ed automaticamente leggibile e modificabile** (p. es., sì alle tabelle editabili, no ai PDF scannerizzati). E il riutilizzo non deve essere solo tecnicamente fattibile, ma anche chiaramente autorizzato dal punto di vista legale. Per questo, **è bene associare ai dati delle licenze/liberatorie, che chiariscano in modo inequivocabile la libertà di riutilizzarli** (e redistribuirli).

Tutti i riusi dei dati dovranno comunque avvenire nel rispetto della legge. Ma, a parte questo ovvio limite, qualsiasi altra restrizione non è che un freno all'innovazione e alla creatività. Nell'ambito della comunità open data c'è sostanziale convergenza verso una sola possibile condizione: l'imposizione di un requisito di attribuzione (o citazione della fonte/origine dei dati). Le piccole incombenze che questo requisito può causare al riutilizzatore, infatti, sono compensate dall'interesse a render merito alle amministrazioni che aprono i loro dati. (*chapeau* a chi non pretenda che gli sia reso merito: tipicamente, gli attivisti dei dati aperti lo loderanno ancor più caldamente!)

Tra i vari errori da evitare, uno merita una (ulteriore) menzione speciale: no alle restrizioni "non-commerciali". **Se il riutilizzo commerciale è vietato, i dati non sono veramente aperti**. Un divieto di riutilizzo commerciale, infatti, sterilizzerebbe la possibilità di molti tipi di riutilizzo creativo da parte di aziende e imprenditori e metterebbe in forse anche la maggior parte dei riusi (di per sé non commerciali) da parte delle *community online* (da Wikipedia a OpenStreetMap), come questo libro spiega in dettaglio.

La conclusione? Mettere a disposizione, a partire da subito e facendo un passo alla volta, i dati grezzi della pubblica amministrazione, così come questa li usa. Poi si possono fare cose meravigliose, ad esempio usando le tecnologie del Web semantico, per rendere i dati ancor più facilmente combinabili e utilizzabili e quasi "intelligenti". Ma il "poi" non deve mettere in discussione il "subito".

RAW DATA NOW. DATI GREZZI, ADESSO!

1

Il valore dell'informazione detenuta dal settore pubblico

1.1 Introduzione: la PSI come opportunità

Le pubbliche amministrazioni (PA) hanno una caratteristica comune: **accumulano dati**. Li accumulano per poter funzionare e per dimostrare di aver ben funzionato (o, se preferite, per controllare che i cittadini rispettino la legge e per poter dimostrare di averla rispettata esse stesse). E lo fanno, si può dire, da sempre: probabilmente, è anche per gestire la contabilità statale che la scrittura si è diffusa in Mesopotamia e in Egitto! Negli ultimi decenni, tuttavia, qualcosa è cambiato nella natura di questa prodigiosa mole di informazione: si tratta, sempre più spesso, di informazioni **in formato digitale** (e, come vedremo, questo ha conseguenze dirompenti, perché rende i dati più facilmente distribuibili e riutilizzabili).

A volte tali informazioni sono generate direttamente come risultato primario dell'attività di un'amministrazione: pensiamo ai dati cartografici, al catasto, alle informazioni meteorologiche, o all'archivio delle emittenti radio-televisive pubbliche. In altri casi, invece, i dati sono acquisiti, organizzati e impiegati dall'amministrazione nell'adempimento dei propri obiettivi istituzionali, ma rappresentano un mezzo rispetto all'output principale – prodotto o servizio che sia. Questo è il caso, per esempio, della banca dati dei tempi di percorrenza medi dei mezzi pubblici in una grande città o della mappa geografica dei codici di avviamento postale.

Tali dati – espliciti e visibili nella prima tipologia di casi sopra delineata (catasto, meteo), e perlopiù invisibili al pubblico nella seconda (perché ad uso interno) – sono nel loro complesso denominati "**informazioni del settore pubblico**" o, in inglese, "**public sector information**" (da cui l'acronimo "**PSI**").

Si noti che la PSI viene generata o raccolta dalle pubbliche amministrazioni in senso stretto, nello svolgimento dei loro compiti istituzionali, ma anche da molti enti pubblici o finanziati da risorse pubbliche. Considerando che, in Italia, le entrate dello Stato sfiorano il 50% del PIL, è ragionevole stimare che anche la proporzione di informazione pubblica rispetto al totale dell'informazione rappresenti una percentuale considerevole.

Quasi tutte le informazioni pubbliche hanno un valore intrinseco, che va oltre l'assolvimento dei compiti istituzionali dell'ente che le detiene. Spesso, la singola informazione ha un'utilità limitata, ma **la raccolta organizzata e la combinazione di più informazioni con fonti differenti può generare un valore aggiunto notevole**. Così, le informazioni geografiche sono alla base di cartine e mappe, le quali possono servire a fornire servizi *online*; mentre i dati catastali possono servire a operatori immobiliari e istituti di credito per fare incontrare la domanda e l'offerta di alloggi e per proporre soluzioni di finanziamento. I metadati relativi

a musei e archivi locali, così come le immagini delle opere in essi disponibili, possono stimolare il turismo. O, ancora, i dati meteorologici possono essere utilizzati sia per servizi a valore aggiunto in tempo reale (ad esempio messaggi inviati sul cellulare dell'abbonato con previsioni selettive del tempo), sia per arricchire il dibattito pubblico sul cambiamento climatico.

In questa prospettiva, l'informazione detenuta dal settore pubblico può essere vista come una grande miniera a cielo aperto, che raccoglie vasti giacimenti di informazioni, dati e contenuti. **Questi dati hanno anche un valore particolare**, che normalmente manca ai dati raccolti dal settore privato: si tratta di dati **raccolti nell'arco di molti anni** e di regola **sottoposti a verifiche** (ad esempio nel corso dell'attività che essi supportano), ma soprattutto si tratta di dati **tendenzialmente "neutrali"** o comunque relativamente affidabili, in quanto meno soggetti a distorsioni generate da interessi commerciali diretti.

Riutilizzo

Chi si occupa di PSI parla spesso di "riutilizzo" (o "riuso"), anziché di "utilizzo" (o "uso") e così facciamo in questo libro. La scelta del termine riutilizzo sottolinea il fatto che stiamo parlando di **usi diversi e ulteriori, rispetto all'uso istituzionale, per cui il dato è stato raccolto o generato dalla PA**.

Il riutilizzo va anche chiaramente distinto dall'accesso. Ciò che ci interessa non è solo la possibilità di accedere ai dati. Approcci tipo "guardare, ma non toccare" non facilitano certo la vita degli sviluppatori e la creazione di servizi utili ai cittadini. Poter vedere i dati è solo il primo passo per poterli acquisire e poi finalmente riutilizzare, cioè modificare, mescolare e/o trasformare per renderli più utili ed interessanti (per alcuni specifici scopi, senza togliere ad altri la possibilità di fare altrettanto, in modo diverso, per i loro differenti obiettivi).

La PSI può essere **riutilizzata** in molte prospettive:

1. nell'ambito di **mercati già esistenti o nuovi, per creare servizi** a valore aggiunto, con benefici di tipo economico. Tornando all'esempio dei dati meteorologici pubblici, non si pensi solo alle classiche previsioni del tempo, ma anche a servizi specializzati, ad esempio per l'agricoltura o il turismo invernale. In alcuni casi, gli introiti fiscali indiretti generati da tali attività possono essere pari o superiori a quelli che la PA otterrebbe facendo direttamente pagare per l'utilizzo dei dati;

2. per contribuire ad un **miglioramento delle scelte pubbliche**, in un quadro di *e-governance* e trasparenza. Ad esempio, il semplice fatto di mettere a disposizione e permettere la rielaborazione di dati quali il numero, il salario o i giorni di assenza degli amministratori e dei dipendenti pubblici può stimolare l'efficienza della PA. In altre parole, condividendo informazioni su sé stessa, **la PA è spinta a migliorarsi**;

3. per consentire forme più sofisticate di **partecipazione dei cittadini** alle scelte pubbliche (*e-democracy*). Ad esempio, la combinazione di dati sanitari e ambientali può permettere a gruppi di persone di valutare l'impatto di diverse scelte politiche e può avere ricadute sulla partecipazione dei cittadini alla vita pubblica. Inoltre, la partecipazione della popolazione può anche passare per la creazione di servizi, simili a quelli che potrebbe offrire lo stato o il mercato, ma su base volontaria e/o associativa.

Vi sono buone ragioni per non lasciarsi sfuggire le opportunità offerte dal riutilizzo dell'informazione del settore pubblico. Non ultima è la circostanza che **i cittadini hanno già pagato, attraverso il prelievo fiscale, per ottenere quello stesso servizio mediante il quale si seleziona, raccoglie o genera il dato**. La PSI può dunque essere messa liberamente a disposizione di chiunque.

Open data

L'espressione open data (dati aperti) o la più specifica open government data (dati pubblici aperti) vengono spesso utilizzate come un sinonimo più accattivante, più cool, di "informazione pubblica" o PSI. PSI finisce così per essere il termine tecnico, usato dai giuristi e dai legislatori europei, mentre open data diventa il termine per la comunicazione informale.

L'uso del termine open data è del tutto accettabile, ma per non ingenerare confusione è necessario che open data sia usato come sinonimo di "informazione pubblica liberamente riutilizzabile" e non di "informazione pubblica" in generale. Anche per questo, nel paragrafo 1.5.1 forniremo una specifica definizione di dati aperti.

Una delle ragioni più significative per rendere disponibili e liberamente riutilizzabili le informazioni detenute dalle PA risiede nella **fantasia dei cittadini e degli imprenditori**. Le PA hanno infatti la disponibilità di enormi quantità di dati, ma hanno anche molti compiti e risorse limitate e spesso gli incentivi e gli strumenti per la creazione o il miglioramento di servizi risultanti dall'elaborazione o aggregazione dei dati disponibili scarseggiano. Per questo, è fondamentale permettere ad altri di riusare l'informazione del settore pubblico. In quest'ottica, **l'informazione del settore pubblico dovrebbe essere vista come una piattaforma**, le cui applicazioni sono ancora quasi tutte da scrivere. L'analogia con Internet come piattaforma di innovazione è calzante. Internet, di per sé, è "solo" una piattaforma, ma ha reso possibile la posta elettronica, il Web e Skype, lo streaming video e centinaia di altre applicazioni, declinate in migliaia di modi differenti. In modo simile, l'iPhone di Apple ha rapidamente catalizzato le capacità creative di migliaia di sviluppatori, i quali lo hanno dotato di innumerevoli applicazioni, molte più di quante Apple avrebbe potuto sviluppare senza aprire (almeno in parte) la propria piattaforma alla creatività del resto del mondo.

1.1.1 Lo scenario internazionale

A livello internazionale, si è ben compreso da tempo che le informazioni pubbliche sono un importantissimo fattore di crescita economica e sociale.

Economica perché sulla PSI è possibile fare imprenditorialità, dalla semplice applicazione per smartphone che fornisce i tempi di arrivo dei mezzi pubblici, a complessi servizi di logistica, ambientali, o altro ancora, e anche perché il riutilizzo della PSI può incrementare l'efficienza sia nel settore privato (se ogni cittadino aspettasse 5 minuti in meno al giorno il tram, quei 5 minuti potrebbero essere spesi per incrementare il PIL, lavorando di più o comprandosi un gelato...), sia all'interno delle PA (che grazie al riutilizzo dei dati possono migliorare anche sé stesse, tramite una migliore interazione con i cittadini e le imprese, per esempio selezionando fornitori più convenienti o focalizzandosi sul miglioramento di servizi essenziali, delegando ai privati o al terzo

settore alcune attività che questi attori possono realizzare autonomamente).

Sociale perché la società civile con la PSI può sia monitorare con maggiore efficacia la PA e la politica (vedere, per esempio, http://www.wheredoesmymoneygo.org), sia sostituirsi ad essa, magari anche solo temporaneamente, laddove essa fosse carente o non sufficientemente agile o veloce (un esempio tra i tanti: il servizio www.seeclickfix.com).

Su queste basi si capisce perché negli USA l'Amministrazione Obama, con uno dei suoi primi atti, abbia disposto la creazione del portale integrato www.data.gov. Un esempio seguito dal governo britannico, con la creazione di www.data.gov.uk, e in seguito in altri paesi, dalla Francia all'Austria. L'Italia sta intraprendendo il medesimo percorso con la piattaforma www.dati.gov.it.

Negli **Stati Uniti**, la creazione di servizi a valore aggiunto ha assunto un'importanza macroeconomica di grande rilievo (più del 7% del PNL); negli ultimi anni, anche per merito dell'Amministrazione Obama, si stanno anche moltiplicando le iniziative di *e-government*.

L'**Europa** ha dovuto fare i conti con alcuni ostacoli che si frappongono all'accesso e al riutilizzo dell'informazione del settore pubblico. Negli USA non c'è copyright sulle informazioni del settore pubblico; la privacy è vista in una prospettiva proprietaria piuttosto che personalistica e viene seguito il principio per cui ai riutilizzatori non può essere addebitata una somma superiore al costo marginale della diffusione dei dati (che spesso è vicino a zero per i dati in formato digitale). In alcuni, se non tutti, gli Stati membri dell'Unione Europea gli atti delle pubbliche amministrazioni sono coperti da copyright o diritto d'autore, ivi compresa l'Italia. A livello Comunitario e di conseguenza nazionale, la privacy è molto rigorosa. Inoltre, in Europa, le divergenze legislative (ad esempio dal punto di vista dell'accesso) rendono molto più difficile la creazione di una dimensione transfrontaliera dell'accesso e del riutilizzo all'informazione del settore pubblico. L'importanza macroeconomica del settore è quindi molto più ridotta (sotto all'1% del PIL, a partire dal 'Rapporto Vickery' divulgato dalla Commissione Europea a dicembre 2011), ma pur sempre rilevante in termini assoluti.

L'Unione europea ha adottato la **Direttiva PSI** (Direttiva 2003/98/CE del Parlamento Europeo e del Consiglio del 17 novembre 2003 relativa al riutilizzo dell'informazione del settore pubblico) al fine di rimuovere alcune fra le barriere che limitano il riutilizzo dell'informazione del settore pubblico. Essa stabilisce alcune **regole minime uniformi**, che gli enti pubblici dovrebbero[1] seguire per rendere disponibili le proprie informazioni e affronta punti fondamentali, come la trasparenza in merito ai dati disponibili e alle condizioni di riutilizzo o la garanzia di non discriminazione o di eque condizioni di concorrenza tra i potenziali utenti. La direttiva è stata recepita nella normativa nazionale in tutti i 27 Stati membri. Anche se la Direttiva PSI non è considerata molto incisiva, poiché lascia pressoché totale margine di discrezionalità agli Stati soprattutto riguardo quali dati rendere effettivamente accessibili (e, dunque, riutilizzabili), l'esistenza di una norma di riferimento a livello Comunitario ha avuto, quantomeno, il pregio di imporre questo tema all'attenzione di tutti i legislatori dell'Unione.

In questo scenario, l'**Italia** ha dato una prima attuazione alla Direttiva PSI con il d. lgs. 24 gennaio 2006, n. 36. Secondo

1 Il condizionale deriva dal fatto che la Direttiva PSI lascia pressoché totale discrezionalità a Stati membri e singole PA rispetto a quali dati mettere a disposizione, nonché una certa discrezionalità rispetto alle modalità di riutilizzo. Non a caso, la direttiva è stata da più parti accusata di avere poco mordente. Vi sono, tuttavia, alcuni paletti minimi di apertura da rispettare, che saranno meglio discussi nel seguito. Inoltre, la Direttiva ha il grande merito di chiarire il favore del legislatore europeo (e, di conseguenza, nazionale) verso le politiche di riutilizzo dei dati.

molti commentatori, era dubitabile che la normativa italiana fosse in linea con la Direttiva PSI, da vari punti di vista. In particolare, essa: faceva salve vecchie norme sui dati catastali e ipotecari; non prevedeva un'informazione su mezzi di ricorso contro il mancato consenso alla riutilizzazione; ammetteva tariffe di riutilizzo determinate con criteri di dubbia compatibilità rispetto al dettato comunitario. Dunque, i primi passi dell'Italia verso l'attuazione della Direttiva PSI sono stati quantomeno timidi. Non a caso, la Commissione ha avviato una procedura di infrazione nei confronti dell'Italia nel marzo del 2009 (interrotta a seguito delle modifiche introdotte dalla legge 96/2010).

Contemporaneamente, a partire dal 2009, il tema open data ha cominciato a diventare mainstream, con alcuni importanti governi (abbiamo menzionato quello americano e inglese, cui va aggiunto almeno quello della **Nuova Zelanda** – si veda www.data.govt.nz) che hanno annunciato e messo in opera importanti iniziative di pubblicazione dei propri dati, con caratteristiche di ampia apertura al riutilizzo. Nel Maggio 2010 è stato lanciato il primo portale open data italiano: il progetto regionale www.dati.piemonte.it. Nel 2011 il tema open data non può ormai più essere considerato di nicchia neppure in Italia, se è vero che il primo semestre dell'anno ha visto probabilmente un'iniziativa sui dati pubblici di interesse potenzialmente nazionale ogni due settimane, compreso un convegno sui dati aperti nella Sala delle Colonne della Camera dei Deputati, nonché numerosi appuntamenti nell'ambito del Forum della Pubblica Amministrazione. Sempre a partire dal 2011, altre pubbliche amministrazioni italiane hanno inaugurato progetti open data.

1.1 2 Le tendenze emergenti...

Il motto delle più avanzate esperienze di apertura dei dati a livello internazionale (spesso di matrice anglosassone) potrebbe essere: "RAW - DATA - NOW - FOR FREE", ovvero, "DATI GREZZI SUBITO E GRATIS". "Grezzi" e "subito" perché i dati scadono, come il latte, e perché i tecnici capaci - dagli ingegneri agli smanettoni - sanno spesso gestirli, anche se non rifiniti e certificati. I dati grezzi, ovvero così come vengono prodotti/raccolti e utilizzati in seno alla PA hanno un duplice vantaggio: sono più facili da rilasciare (in quanto si trovano già in questa condizione dentro la PA stessa) e incorporano il maggior numero di informazioni possibili a tutto vantaggio dei riutilizzatori (i dati elaborati e sintetizzati – l'opposto dei dati grezzi – possono infatti essere più direttamente comprensibili al profano, ma è difficile elaborarli e incrociarli ulteriormente, il che è proprio l'obiettivo del riutilizzo avanzato degli open data).

"Gratis" perché bisogna resistere alla tentazione, fortissima in tempi di ristrettezze di bilancio, di indurre la PA a far pagare i propri dati: è un dato di fatto ormai dimostrato, infatti, che i modesti introiti per la singola PA – ammesso di non essere già annullati dai costi relativi all'incasso (ossia legati alla gestione delle transazioni) – si otterrebbero a scapito degli introiti di fiscalità generale molto più ingenti che si avrebbero se i dati fossero disponibili online gratuitamente. Secondo alcuni studi relativi, per esempio, ai dati meteorologici, l'indotto economico generato rilasciando i dati liberamente online (come viene fatto negli USA) è da *tre* a *dieci* volte superiore a quello generato vendendo i dati a prezzi di mercato (come capita spesso in Europa).

Interscambio

> Questo libro si focalizza sul riutilizzo dei dati pubblici da parte di soggetti privati. **Il tema della condivisione di dati tra diverse pubbliche amministrazioni** – spesso detto "interscambio" – meriterebbe un autonomo approfondimento: qui, ci limitiamo a menzionarlo, al fine di far percepire ai decisori pubblici anche questa opportunità (alla quale si applicano comunque alcuni dei discorsi che faremo in questo lavoro, ad esempio quelli relativi ai formati aperti e standard).

1.1 3 ...e gli errori da evitare

Tuttavia, e malgrado una crescente consapevolezza presso alcuni dei gruppi più dinamici all'interno della PA italiana a livello sia centrale sia di autonomie locali, l'Italia non ha ancora sposato in modo deciso la filosofia dei dati aperti.

Spesso, **manca la consapevolezza che le pubbliche amministrazioni si possano sempre più considerare anche (e, in alcuni casi, soprattutto) come entità che acquisiscono, producono, elaborano e comunicano informazione**. Senza tale consapevolezza, è quasi normale che i dati vengano di norma prodotti e usati al bisogno, a seconda delle specifiche circostanze, ma senza linee guida omogenee, senza tener conto di usi e riusi diversi da quelli per i quali vengono originariamente prodotti, senza una visione né d'insieme né di lungo termine. **Chi questa consapevolezza ce l'ha, invece, rischia di essere vittima di una visione proprietaria dell'informazione pubblica**. Non a caso, questo libro bianco si propone di impostare una specifica "narrativa" dell'informazione del settore pubblico, la quale **non deve essere percepita come proprietà dell'ente, ma come patrimonio dei cittadini**. Infatti, per quanto gli enti pubblici investano milioni per raccogliere, gestire e mantenere i propri database, l'investimento fatto dall'ente è a sua volta **frutto di tasse pagate dai cittadini** (e dalle imprese); inoltre – e a volte questo è quel che più conta – i dati stessi sono normalmente **creati dai cittadini stessi** nell'adempiere ad obblighi di vario genere e spendendo tempo e denaro. In un simile scenario, il valore aggiunto del ruolo pubblico potrebbe essere quello di restituire agli stessi cittadini un insieme di dati verificati e organizzati. E – in una sorta di versione digitale della parabola dei talenti – **il peggior errore sarebbe quello di chiudere in un cassetto il patrimonio informativo ricevuto in custodia dai cittadini**.

Le diverse forze in gioco generano **un insieme di informazioni del settore pubblico altamente frammentato e variabile**. I dati possono essere disponibili (a volte) o meno (assai più spesso). Possono essere disponibili a tutti (di rado) o solo a favore di determinate entità (troppo spesso). La loro stessa esistenza può essere nota grazie ad appositi registri (quasi mai), o invece essere ignota al pubblico o ad altre amministrazioni o persino ad altri uffici della medesima amministrazione (in alcuni casi per difficoltà oggettive, ma molto più spesso per eccessiva, e a volte immotivata, prudenza). Possono essere disponibili a prezzo di mercato (o quasi di monopolio), oppure al costo marginale (ovvero al costo di distribuire una copia in più, legato alla banda necessaria o alla riproduzione di un supporto, p. es. un DVD), o gratis. Possono essere preservati per anni dall'entità che li ha prodotti oppure scartati subito dopo l'uso. Possono avere qualità garantita o meno. Possono essere resi disponibili con formati aperti e ben documentati, o con formati che vincolano all'uso di determinati software o apparati.

Possono venire utilizzati per qualsiasi scopo (per esempio, anche commerciale) o meno. E così via.

Il prezzo che già oggi paghiamo per l'assenza di strategia e di iniziative in merito alle informazioni del settore pubblico è enorme e destinato a crescere ulteriormente. Una nuova consapevolezza è dunque necessaria e questo libro bianco nasce proprio come un modesto contributo a generarla e a favorire l'individuazione e la diffusione di buone pratiche relative alla gestione del patrimonio informativo della collettività, in ottica open data.

1.2 La PSI come piattaforma di innovazione (tecnologica e sociale)

In questo paragrafo forniremo alcuni esempi di come l'accesso alla e il riutilizzo della PSI siano un'opportunità di innovazione e guadagno per le imprese (e dunque di crescita e tasse per i governi), nonché di crescita culturale e sociale per i cittadini (tramite maggiore partecipazione alla vita culturale, democratica e sociale in generale). Cercheremo anche di mostrare come la stessa PA possa cogliere questa sfida in ottica innovativa. Varie mode, più o meno passeggere, hanno portato all'ordine del giorno il tema *e-government*, *government 2.0* e, più di recente, open government. Comunque lo si chiami, l'idea è sempre la stessa e potenzialmente ottima. Favorire una maggiore partecipazione dei cittadini alle scelte della PA, renderli più partecipi e far sì che l'amministrazione possa imparare da loro (dai loro bisogni, dalle loro percezioni, ma anche dalle loro competenze tecniche o semplicemente dalle loro idee che nascono da un punto di vista differente). In altre parole, si tratta di governare valorizzando il meglio del lavoro della PA e facilitando l'eliminazione della parti peggiori (anche tramite la loro evoluzione); tutto ciò, potenzialmente migliorando la percezione di alcuni adempimenti oggi vissuti da cittadini e imprese (a torto o a ragione) come pura vessazione burocratica. Infatti, sapere che una medicina un po' amara fa bene può anche renderla meno sgradevole da ingoiare e, per la stessa ragione, i cittadini che hanno gli strumenti per capire e controllare il lavoro della PA sono spesso meno propensi a considerarlo inutile o ingiusto.

A questo punto, vi starete probabilmente chiedendo: "Le amministrazioni hanno investito un sacco di soldi in progetti di *e-government*, li stanno ancora investendo in servizi interattivi di *government 2.0*: cambio di etichetta a parte, perché dovremmo entusiasmarci per l'*open government* e l'open data?". Probabilmente, la risposta più convincente è: per il cambiamento di prospettiva. L'*e-government* si proponeva di generare miglioramenti per i cittadini tramite l'introduzione di strumenti digitali, spesso informatizzando i tradizionali schemi di interazione tra amministrazioni e cittadini. Il government 2.0 innova questo schema sottolineando il ruolo centrale dell'interazione coi cittadini/utenti e degli strumenti di *social networking*. L'*open government*, invece, riguarda meno la tecnologia (meglio: una specifica tecnologia) e molto di più il rapporto tra governanti e governati, amministratori e cittadini. In particolare, anziché focalizzarsi sull'offerta di specifici servizi, si mettono i cittadini in grado di capire e partecipare, dotandosi essi stessi di nuovi strumenti. Focalizzandoci sul tema open (government) data (che non coincide con l'open government, ma che ne è un aspetto caratterizzante), la rivoluzione copernicana sta nel fornire al cittadino non un nuovo mezzo tecnologico per accedere ai servizi pubblici, ma **i dati necessari a supportare, affiancare o criticare l'operato dell'amministrazione pubblica stessa**. Al limite anche

fornendo un servizio concorrente, se – con gli stessi dati – qualcuno pensa di poter far meglio della pubblica amministrazione[2]. La PSI diventa così una piattaforma informativa, fertile rispetto a qualsiasi tipo di innovazioni. L'idea è del tipo "invece del pesce, ti offro il necessario a costruirti una canna da pesca – e la mappa per raggiungere il fiume" (naturalmente – continuando con la metafora – ti fornisco anche il permesso di pesca e stabilisco alcune regole, per evitare che qualcuno pensi di poter andare a pescare con la dinamite). Uno dei principali vantaggi di questo modello è che più approcci alternativi diventano possibili, perché i cittadini (e le loro associazioni) e gli imprenditori possono sperimentare. Ed è più difficile sbagliare strada e, ad esempio, investire risorse pubbliche in un servizio, che potrebbe essere molto utile, ma che nessuno è ancora pronto ad utilizzare.

Perché questo approccio open data possa funzionare, però, è necessario rinunciare al controllo totale. La PSI nasce dentro le pubbliche amministrazioni e può crescere, diventano fonte di innovazione per i cittadini, le imprese e la stessa PA. Tuttavia, come tutti i figli, per crescere bene ad un certo punto la PSI deve abbandonare la casa materna. Ovviamente – nelle storie che più valgono la pena di essere raccontate – ci saranno occasioni di ritorno all'ovile, perché l'informazione pubblica, usata dai privati, genererà nuovi servizi pubblici e aiuterà la stessa pubblica amministrazione a migliorarsi. Ma, se la pubblica amministrazione pretende di controllare, prevedere, capire in anticipo tutto quello che ognuno potrà fare con le informazioni pubbliche, molto difficilmente queste potranno essere riutilizzate in quei modi inattesi, che hanno la più alta probabilità di creare benefici per tutti.

I dati detenuti dalle amministrazioni possono, molto semplicemente, essere utili in modo diretto ai cittadini, se qualcuno si fa carico di renderli più accessibili e *user friendly*: ad esempio, una donna danese ha realizzato findtoilet.dk, un servizio per individuare bagni pubblici, prendendo spunto dalle necessità di suoi conoscenti con problemi di incontinenza.

Un campo complesso è quello della trasparenza. Un possibile scopo della trasparenza è il controllo dell'operato della PA. Maggiore trasparenza, però, può anche voler dire maggior coinvolgimento e maggiore partecipazione: sapere dove finiscono i soldi delle mie tasse, per esempio, può renderle il pagarle più accettabile. In questo senso, è interessante il progetto britannico Wheredoesmymoneygo.org ("dove finiscono i miei soldi")[3]. Secondo gli sviluppatori del servizio, molti utilizzatori sono rimasti favorevolmente sorpresi dall'importanza della macro voce di spesa "aiutare gli altri", che raccoglie ad esempio le spese di assistenza ai soggetti più deboli della società. Allo stesso modo, alcuni sono rimasti stupiti dalla relativa esiguità di voci di spesa spesso criticate, come quelle militari.

Naturalmente, mescolando più fonti di informazioni differenti si possono aiutare i cittadini a prendere decisioni più complesse. Ad esempio, Mapumental (relativo al Regno Unito: www.mapumental.com) e Mapnificent (creato in Germania, ma che fa uso di dati da tutto il mondo: www.

2 A scanso di equivoci, in un caso del genere, non è detto che la pubblica amministrazione debba far venir meno il proprio servizio — sono molti i campi in cui la coesistenza di un servizio pubblico e di più servizi privati, alcuni dei quali imprenditoriali e altri non-profit, risulta socialmente desiderabile, poiché ciascuno ha i propri pregi e i propri difetti, spesso mitigati dalla presenza di alternative. Ad esempio, un'azienda municipalizzata vorrà mettere a disposizione sul proprio sito un servizio accessibile via Web per calcolare percorsi urbani coi mezzi pubblici, ma sarà opportuno facilitare il lavoro di chi voglia realizzare un'App per l'iPhone o per telefoni Android che faccia la stessa cosa, in modo ottimizzato per queste piattaforme software.

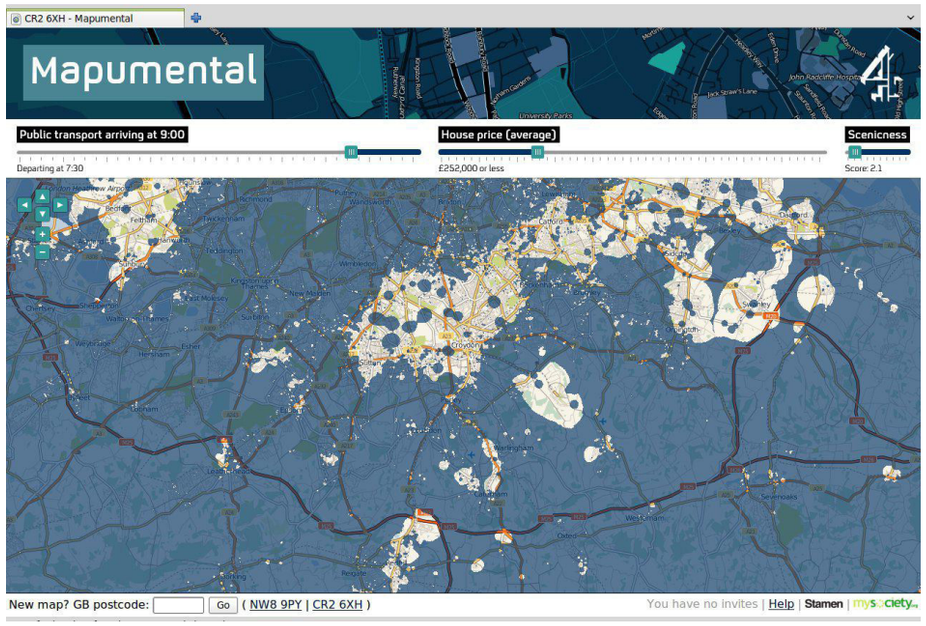
3 L'attuale evoluzione internazionale del progetto è http://www.openspending.org/, che tra l'altro ha visto alcune informazioni sull'Italia tra i primi dataset internazionali ospitati: http://www.openspending.org/dataset/italyregionalaccounts.

mapnificient.net) aiutano i cittadini a scegliere l'area in cui stabilirsi, considerando il tempo necessario a raggiungere il proprio luogo di lavoro con i mezzi pubblici, il prezzo delle case e altri indicatori relativi alla qualità della vita nella zona in questione.

Mapumental

Mapumental è una applicazione web che offre agli utenti una mappa interattiva, combinando tra loro dati geografici, dati catastali, dati dei servizi di trasporto pubblico e dati generati dagli utenti del Web. Sviluppato da mySociety congiuntamente a Channel 4's 4IP, questo strumento permette di individuare il luogo ideale della Gran Bretagna in cui stabilirsi, in base alle proprie esigenze di tempo negli spostamenti, possibilità economiche e aspettative circa la gradevolezza urbanistica ("scenograficità"). Selezionando un codice di avviamento postale corrispondente ad una destinazione di riferimento (ad esempio, il proprio luogo di lavoro abituale), e scorrendo le barre relative al tempo che si è disposti ad impiegare per raggiungerla entro le ore 9:00, al costo degli immobili e alla scenograficità del quartiere, è possibile circoscrivere e visualizzare le zone corrispondenti ai parametri scelti, individuando la propria dimora ideale.

Mapumental si basa, incrociandoli tra loro, su una serie di dati di provenienza diversa. **Le mappe di sfondo** sono generate sfruttando OpenStreetMap, che a sua volta riusa informazioni geografiche pubbliche, oltre a quelle generate dagli utenti. **I dati del trasporto pubblico** provengono dalla banca dati National Public Transport Data Repository (NPTDR). **I prezzi degli immobili** sono stati ricavati acquistando una lista di tutte le vendite di proprietà immobiliari ad uso abitativo registrate presso il Land Registry di Inghilterra e Galles tra gennaio 2008 e marzo 2009 (i dati relativi alla Scozia non sono stati inclusi perché troppo costosi). **La "scenograficità"** è stata stimata sfruttando le valutazioni sulla gradevolezza estetica di diverse aree delll'Inghilterra, della Scozia e del Galles, espresse dagli utenti del sito ScenicOrNot, creato appositamente per accumulare i dati necessari per Mapumental. Le foto presenti su ScenicOrNot provengono dal sito Geograph.



Come abbiamo cercato di mostrare in queste pagine, esistono numerosi esempi concreti e già funzionanti di riutilizzo dell'informazione pubblica. Tuttavia, bisogna tenere a mente che questa è solo la punta di un iceberg. Il processo di apertura dei dati pubblici è ancora agli inizi: molti più usi creativi saranno possibili al crescere del numero dei dati (poiché crescerà ancor più rapidamente il numero

di combinazioni possibili). E la crescente alfabetizzazione informatica dei cittadini, unita alla disponibilità di software sempre più intuitivi, aumenterà anche il numero di riutilizzatori potenziali.

1.3 La tecnologia

PSI ante litteram

> Nel 1854 a Londra si diffuse un'epidemia di colera. Il dottor Snow, cercando di comprenderne le cause, utilizzò una piantina di Londra (un tipico esempio di informazione normalmente raccolta dal settore pubblico), riportandovi la diffusione dei casi nei diversi periodi (un'altra informazione potenzialmente disponibile al settore pubblico). Questo metodo gli permise di notare come i casi si concentrassero in corrispondenza di una pompa dell'acqua nella zona di Soho, che Snow intuì essere inquinata dagli scarichi delle abitazioni, ancora prive di un vero sistema fognario. Bloccata la pompa, il diffondersi della malattia tornò sotto controllo. Purtroppo, ci vollero diversi anni prima che l'intuizione di Snow (e la sua teoria relativa alla causa batteriologica del colera) venisse universalmente riconosciuta e un sistema fognario appropriato garantisse contro l'inquinamento delle acque potabili londinesi... ciò nonostante, il caso delle ricerche di Snow rappresenta un esempio di riutilizzo dell'informazione pubblica, con un immediato ritorno sulle decisioni di policy (la chiusura, sia pure temporanea, della pompa dell'acqua) e un diretto beneficio per la società.

Come abbiamo accennato sin dalle prime righe di questo lavoro, le amministrazioni pubbliche raccolgono grandi quantità di dati praticamente da sempre. Se i dati pubblici ci sono sempre stati, e in abbondanza, **è naturale chiedersi perché il riutilizzo della PSI sia un problema da affrontarsi proprio ora**. In altre parole, tutta questa attenzione relativa al riutilizzo delle informazioni non sarà mica una moda?

Ognuno può darsi la propria risposta, ma noi pensiamo che la tendenza verso una maggiore condivisione dei dati pubblici sia una processo molto serio e probabilmente inarrestabile: non certo una moda passeggera. Se in passato non si parlava molto di dati pubblici aperti (o si parlava solo di diritto d'accesso, in ottica di trasparenza) è perché ci troviamo oggi in una situazione nuova. **Le ragioni che rendono opportuno aprire e condividere i dati – proprio adesso – sono fondamentalmente legate a sviluppi tecnologici**. Oggi, infatti, condividere dati digitali costa poco, quasi nulla. Ed elaborarli è facile e veloce. In questo scenario, diventano possibili – anche solo nella logica del "Perché no?" – cose che sino a qualche anno or sono erano semplicemente infattibili o così costose da essere proibitive.

Come speriamo alcuni lettori stiano già sospettando, **un'argomentazione forte a favore dell'apertura dei dati pubblici è proprio quella del "Perché no?"**. Perché, infatti, si dovrebbe sottrarre alla creatività, alla passione, allo spirito d'iniziativa, alla fantasia e alle tante altre risorse dei cittadini e degli imprenditori una ricca dote, che la PA ha comunque a propria disposizione, per svolgere alcune sue funzioni necessarie, ma che non riesce a riutilizzare produttivamente, se non in minima parte. Perché no, se il costo è ridotto? Perché no, se l'esperienza insegna che gli utenti della Rete riescono a portare avanti progetti ambiziosi del calibro di un'enciclopedia online come Wikipedia? La tecnologia, infatti, non ha solo ridotto i costi per distribuire e processare i dati, ha anche messo a disposizione delle

persone piattaforme adatte ad esercitare forme di **collaborazione di massa**, capaci di generare una sorta di intelligenza collettiva.

Diamo i numeri!

Quanto vale l'informazione del settore pubblico e quanto costa metterla a disposizione? Tentando di rispondere a questa domanda si rischia veramente di "dare i numeri", poiché le stime variano significativamente. Tra gli studi internazionali sul valore potenziale della PSI, uno dei primi è il cosiddetto studio Pira (2000), che parla di un valore di 750 miliardi di dollari negli USA, contro 68 miliardi di Euro nella UE. Differenti metodologie e approcci hanno portato il successivo studio MEPSIR (2006) a parlare di 27 miliardi di Euro nella UE. All'incirca nello stesso periodo, l'Office of Fair Trading (OFT) del Regno Unito ha compiuto un proprio studio, che parla di un valore di 600 milioni di sterline nel Regno Unito. Operando alcuni correttivi per rendere comparabili i vari studi, la società Dialogic ha suggerito stime che si aggirano sui 5 miliardi di Euro come valore diretto (per la UE) e sui 30 miliardi, considerando l'indotto. Una recente analisi della Commissione Europea stima a **140 miliardi di euro all'anno l'impatto diretto e indiretto delle applicazioni basate su PSI** nel complesso dell'Unione.

Alcuni considerano queste stime faraoniche; altri fanno notare che, in realtà, sono probabilmente troppo esigue, perché non considerano i fondamentali effetti indiretti di aumento di efficienza per la PA e per i cittadini stessi. Fortunatamente, che uno voglia o meno credere a questi numeri, realizzare un progetto per mettere a disposizione del pubblico la PSI in forma riusabile sembra resti una buona idea. Infatti, **stime approssimative del costo dei vari portali nazionali e regionali esistenti in Europa si aggirano nell'ordine dei 150-300 mila Euro** (il che pare coerente con una recente call for tender della Commissione Europea per il proprio portale open data, il cui costo non dovrà superare gli 800 mila Euro). Ovviamente, a questi costi si devono aggiungere un po' di mesi/uomo destinati alla parte informale del processo di apertura dei dati, ma sembra si possa ragionevolmente sostenere che **l'ordine di grandezza dei benefici è circa mille volte maggiore di quello dei costi**. Dunque, perché no?

Uno dei principi fondamentali della *new economy*, che sembra applicarsi bene anche all'approccio open data è che **"l'intelligenza è là fuori": ci sarà sempre (o quasi) qualcuno capace di usare i dati che custodite in modi che voi non avreste potuto immaginare**. E questo non perché la PA non abbia persone capaci e volenterose al proprio interno, ma semplicemente perché le persone all'esterno di ogni organizzazione – dal Comune di Moncenisio a Google – sono sempre molte di più, hanno interessi disparati e punti di vista differenti. E possono collaborare tra loro utilizzando le piattaforme più varie.

PSI e contenuti generati dagli utenti

Oltre ad abbassare il costo del riutilizzo, la tecnologica offre nuove piattaforme e nuovi contesti sociali per metterlo in pratica. Infatti, non solo gli utenti di Internet rappresentano la principale fonte di domanda per i dati del settore pubblico e per le applicazioni basate su di essi. **I contenuti generati dagli**

utenti rappresentano anche il miglior complemento rispetto all'informazione pubblica ed il Web offre la possibilità di aggregare e organizzare questi contributi, combinandoli a loro volta con l'informazione pubblica. Ad esempio, una carta geografica basata su dati pubblici può essere arricchita dalle segnalazioni e dalle foto degli utenti (come individui o come professionisti di vario genere, p. es. albergatori), come Google ha chiaramente mostrato con il proprio servizio GoogleMaps. E, sempre a partire da carte geografiche pubbliche, si può anche creare una piattaforma che chieda e permetta agli utenti di caricare foto (magari geolocalizzate tramite il loro smarphone), abbinate a report su problemi quali le buche nelle strade o gli arredi urbani danneggiati e si può anche **creare un canale di ritorno**, dalle pubbliche amministrazioni che prendono in carico queste segnalazioni ai cittadini: così ha fatto www.fixmystreet.com.

Public sector content

Oltre a generare e raccogliere grandi moli di informazione, il settore pubblico è custode di grandi quantità di oggetti, dalle quali si possono ricavare enormi quantità di informazioni (che attualmente sono ad uno stato così grezzo da non rappresentare vere informazioni, ma solo oggetti, artefatti). Pensate, per esempio, alle collezioni di musei e archivi. Questo materiale è a volte chiamato "contenuti detenuti dal settore pubblico" (o *public sector content*). Anche limitandosi agli oggetti che non sono mai stati protetti dal diritto d'autore (p. es. una statua romana) o che non lo sono più perché antichi e ormai caduti in pubblico dominio (p. es. un libro del Settecento), stiamo parlando di milioni e milioni di artefatti, ciascuno dei quali è una miniera di informazioni (che le amministrazioni pubbliche custodiscono, ma che spesso non hanno le risorse per valorizzare).

Nel momento in cui parte di questi "contenuti detenuti dal settore pubblico" vengono digitalizzati (e i progetti di digitalizzazione meriterebbero, di per sé, un libro bianco), ci troviamo di fronte ad un'informazione digitale, che può essere considerata a tutti gli effetti PSI, così come è PSI l'insieme dei metadati relativi a questi contenuti (p. es., il catalogo digitale di una biblioteca pubblica). Purtroppo, in questo libro bianco non abbiamo lo spazio necessario ad approfondire questo tema, ma speriamo comunque di aver suscitato la vostra curiosità!

1.4 L'accessibilità dei dati

Questo libro bianco si focalizza sul riutilizzo dei dati. Ovviamente, non ci può essere riutilizzo se i dati non sono accessibili. Tuttavia, siccome speriamo che i nostri lettori siano amministratori di buona volontà, il nostro auspicio è che illustrare i benefici del riutilizzo – ed accennare, come faremo nella parte finale di questo lavoro, ad alcuni aspetti delicati, dalla proprietà intellettuale alla privacy – possa fornire uno stimolo sufficiente – e i *caveat* necessari – anche rispetto all'offrire libero accesso ad una quantità crescente di dati. In quest'ottica, ci focalizziamo qui sui principi generali che possono essere invocati a proprio sostegno da un amministratore che voglia legittimare un'ampia messa a disposizione dei dati pubblici.

Il Decreto legislativo 24.1.2006, n. 36 (di "Attuazione della direttiva 2003/98/CE relativa ai riutilizzo di documenti nel settore pubblico") prevede (art. 1, comma 2) che "[l]a decisione di consentire o meno tale riutilizzo spetta all'amministrazione o all'organismo interessato, salvo diversa previsione di legge o regolamento". Salvo norme specifiche (e vedremo sotto il caso dei dati ambientali), vi è dunque un ampio margine di discrezionalità. Questa libertà di manovra, tuttavia, è limitata da alcuni principi fondamentali, a livello internazionale, che devono essere rispettati da tutte le leggi nazionali.

Il diritto all'informazione come diritto dell'uomo

Il diritto all'informazione come diritto dell'uomo trova tutela nella Convenzione Europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali (art. 8 e 10) e nella Carta dei diritti fondamentali dell'Unione Europea (art. 7, 8 e 11). Inoltre, nell'ambito del diritto italiano, il diritto all'accesso alle informazioni per tutti i cittadini deve essere ricondotto all'art. 21 della Carta Costituzionale, nonché agli articoli art. 97 e 98. Si evidenzia dunque un contesto favorevole per le pubbliche amministrazioni che vogliano consentire l'accesso ai dati in loro possesso. E, in certi casi, può configurarsi un esplicito diritto del cittadino alle informazioni, con conseguente obbligo per l'amministrazione a metterli a disposizione.

Oltre alle norme internazionali, può contare il buon esempio, fornito anche dalle istituzioni europee.

L'accesso ai dati delle istituzioni europee

Presente sin dai primi trattati comunitari, attualmente la norma di riferimento in tema di accesso ai dati è il capo V della Carta dei diritti fondamentali, dedicato ai diritti di cittadinanza, dove viene espressamente previsto il diritto di accesso ai documenti: "Ogni cittadino dell'Unione nonché ogni persona fisica o giuridica che risieda o abbia la sede sociale in uno Stato membro ha il diritto di accedere ai documenti delle istituzioni, organi e organismi dell'Unione, a prescindere dal loro supporto" (art. 42 della Carta). Il diritto di accesso agli atti, viene dunque a godere di un doppio statuto: quello di diritto fondamentale e quello di principio generale di diritto dell'Unione europea.

Il nuovo articolo 15 del TFUE, stabilisce anche che "[a]l fine di promuovere il buon governo e garantire la partecipazione della società civile, le istituzioni, gli organi e gli organismi dell'Unione operano nel modo più trasparente possibile." Il diritto di accesso viene quindi affermato sia quale principio generale di trasparenza dell'azione dell'Unione Europea, sia quale strumento per promuovere il buon governo (c.d. *good governance*) e garantire la partecipazione della società civile.

Il diritto di accesso agli atti delle istituzioni dell'Unione europea, a seguito di una trentennale evoluzione, è ricostruibile come un vero e proprio diritto del cittadino, azionabile avanti la Corte di giustizia. Infatti, l'accesso ai documenti non è preordinato alla tutela di una propria posizione giuridica soggettiva, quindi non richiede la prova di un interesse specifico, ma risponde ad un principio generale di trasparenza dell'azione dell'Unione ed è uno strumento di controllo democratico sull'operato

dell'amministrazione europea. L'evoluzione dell'accesso ai documenti come principio generale dell'azione delle istituzioni e come diritto fondamentale del cittadino può considerarsi compiuta.

Date le rilevanti interazioni col tema dell'accesso, anticipiamo qui che la normativa sulla protezione dei dati personali (sulla "pryvac", per brevità) rappresenta un limite importante all'informazione che può essere resa pubblicamente accessibile (e anche al tipo di riutilizzo che si può fare di alcuni tipi di informazioni personali, raccolte sulla base di un ben preciso e dichiarato scopo, come la legge sulla privacy prevede).

Dal punto di vista di chi voglia massimizzare le opportunità di riutilizzo, la privacy può a volte sembrare un ostacolo – ed in effetti capita che sia usata pretestuosamente per evitare di pubblicare dati più "scomodi" per qualcuno che "sensibili" dal punto di vista della legge sulla tutela dei dati. Ciò premesso, è bene tenere a mente che **il rispetto dei dati personali degli individui è un elemento importante, necessario a salvaguardare la fiducia dei cittadini verso le istituzioni**. È dunque ovvio (e sancito dalla legge, a scanso di equivoci) che l'obiettivo dell'apertura dei dati e del loro riutilizzo non giustifichi la violazione del diritto alla riservatezza.

Come nota finale di questa sezione, menzioniamo comunque l'esistenza di un insieme di norme relative all'accesso ai dati, dal punto di vista di un cittadino, associazione o impresa, che intenda formulare una domanda di accesso, eventualmente anche contro la volontà dell'amministrazione che detiene i dati stessi. In questo campo, purtroppo, non esiste una normativa uniforme a livello Europeo e l'Italia non brilla, rispetto a paesi dove il *Freedom of Information Act* è un potente strumento di trasparenza. Ciò detto, l'accesso ai documenti amministrativi è regolato dalla legge 241/90, cui rimandiamo per ulteriori approfondimenti. Per una disanima maggiormente esaustiva, si segnala il lavoro di Francesca Pavoni e Paolo Patrito, disponibile al seguente link: http://www.evpsi.org/evpsifiles/fpavonippatritoEVPSI2ndolivello.pdf.

Accesso alle informazioni & tutela dell'ambiente

La Convenzione di Aarhus, trattato internazionale in materia ambientale del 25 giugno 1998, ha sancito il collegamento tra la protezione dell'ambiente ed i diritti umani. Gli Stati membri devono garantire agli individui la possibilità di essere coinvolti nel processo di gestione dell'ambiente, mediante l'accesso all'informazione, la partecipazione ai processi decisionali e l'accesso alla giustizia. Il diritto d'accesso all'informazione nella Convenzione di Aarhus viene garantito sia sul versante passivo, l'obbligo di fornire le specifiche informazioni richieste dai cittadini (senza che il richiedente debba vantare un interesse particolare), sia sul versante attivo, l'obbligo per le autorità pubbliche di raccogliere e divulgare l'informazione in materia ambientale a prescindere da qualsiasi richiesta.

Sul versante dell'Unione Europea, sono particolarmente rilevanti la direttiva 2003/4/CE, sull'accesso del pubblico all'informazione ambientale e la direttiva 2003/35/CE, relativa alla partecipazione del pubblico al processo decisionale.

In Italia, l'accesso alle informazioni ambientali si distingue dal generico accesso agli atti (legge 241/90) perché non necessita l'individuazione di un interesse

meritevole di tutela al fini dell'accesso stesso. In un certo senso, l'ordinamento italiano ha addirittura anticipato la normativa internazionale ed europea in materia di accesso alle informazioni ambientali, con la legge 349/1986 istitutiva del Ministero dell'Ambiente. Più recentemente, il D.Lgs. 195/2005, inoltre, recepisce la Direttiva 2003/4/CE nell'ordinamento italiano.

1.5 Una (prima) raccomandazione: "Dati grezzi aperti subito!"

Quanto detto sinora ci spinge ad ipotizzare una prima possibile raccomandazione: "Dati grezzi aperti subito!". Rispetto allo slogan "Raw data now!", ci permettiamo di esplicitare l'importanza che i dati siano aperti. Questo perché non basta che i dati siano praticamente disponibili e tecnicamente riutilizzabili; vogliamo anche che siano riutilizzabili legalmente.

1.5.1 Dati aperti

È ormai tempo di esplicitare cosa intendiamo esattamente per "aperto" nel campo dei dati o della conoscenza. Per i nostri fini si può adottare una definizione, che deriva dall'Open Definition (http://opendefinition.org/): **i dati aperti sono dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque – soggetti solamente, al più, al requisito di attribuzione e condivisione nello stesso modo**.

Da questa definizione discende che i dati sono aperti solo se si applica una delle seguenti condizioni: sono automaticamente liberi da qualsiasi tipo di diritto di proprietà intellettuale (ovvero sono **in "pubblico dominio" per legge**); vengono messi a disposizione sotto licenze/liberatorie, che escludono qualsiasi limitazione (ovvero vengono posti contrattualmente in una situazione equivalente al pubblico dominio), come la licenza/liberatoria **Creative Commons Zero (CC0)**; vengono licenziati a condizioni molto permissive, che richiedano solo l'attribuzione/citazione della fonte, per esempio tramite la licenza **Creative Commons Attribuzione (CC BY)**, la **Italian Open Data License v. 2.0 (IOLD 2.0)** o la licenza **Open Data Commons Attribution (ODC BY)**; vengono licenziati sotto licenze permissive, ma con clausole share-alike, per esempio tramite la licenza Creative Commons Attribuzione Condividi-allo-stesso-modo (CC BY-SA) o Open Data Commons Open Database License (ODC OdbL). Per ulteriori approfondimenti, rimandiamo al paragrafo 2.3.

Un altro importante corollario della definizione di dati aperti adottata in questo libro bianco (e dalla comunità open data a livello internazionale) è che non è consentita nessuna discriminazione dei riutilizzatori o delle classi di riutilizzi. In particolare, **le restrizioni di tipo "non-commerciale" non sono ammesse**. Ovvero, **se il riutilizzo commerciale è vietato, i dati non sono veramente aperti**. Questo è un punto fondamentale[4]. Ed è anche un punto che farà storcere sicuramente il naso a qualcuno. Riflettete tuttavia sull'effetto di un divieto di riutilizzo commerciale. In primo luogo, si sterilizzerebbe la possibilità di molti tipi di riutilizzo creativo

4 È dunque chiaro che non tutta la PSI riutilizzabile è anche aperta... per esempio, riguardo le restrizioni al riutilizzo commerciale, ci sono molti distinguo e il legislatore italiano, nel suo recepimento della Direttiva PSI, lascia aperta la porta a licenze che discrimino tra riusi commerciali e non commerciali.

da parte di aziende e imprenditori individuali. Inoltre, si metterebbero in forse anche molti modelli di riutilizzo che puntano semplicemente a rientrare dei costi (magari dei costi medi e non del costo netto di ogni nuovo accesso al servizio innovativo creato). Infine – e questa è forse la cosa più grave – **le clausole non-commerciali limitano la catena dei riusi** e, in particolare, rendono impossibile la mescolanza coi dati generati da progetti come Wikipedia, dbPedia e OpenStreetMap, i quali sono chiaramente progetti non commerciali, che lasciano tuttavia ampia libertà di riusare i dati che generano, anche a chi voglia trarne un guadagno. Questa **incompatibilità di ogni clausola non-commerciale con la maggior parte dei progetti di *community online*** deriva dal fatto che tali progetti usano licenze di tipo condividi-allo-stesso-modo, per cui i dati così licenziati possono essere mescolati solo con dati licenziati allo stesso identico modo o – ancor meglio – con dati ancor più liberi, ovvero in pubblico dominio o vincolati alla sola attribuzione della fonte.

2 Apriamo un po' di dati

L'esperienza dei progetti sui dati aperti di maggiore successo, dagli USA alla Nuova Zelanda, passando per il Regno Unito e – in casa nostra – per il Piemonte, ci spinge a fare alcune raccomandazioni, che riassumiamo in una serie di principi. Come sempre, è possibile immaginare molte eccezioni, ma – specialmente qualora non abbiate particolare esperienza nel campo open data – il nostro consiglio è di trattare queste eccezioni come tali: motivatele una a ad una e chiedetevi perché il principio generale non si applichi al vostro caso.

2.1 Le migliori pratiche

Create una squadra

Per poter partire, è importante che ci sia un *leader*, **un politico o un funzionario che ci metta la faccia**. Se questo *leader* non è un alto dirigente, può sempre provare a convincerne uno, affinché lo lasci fare (in fondo, è spesso così che si diventa "eroi dei dati aperti"!). Ma, soprattutto, serve una squadra, che abbia alcune competenze chiave (alcune delle quali possono essere possedute dalla stessa persona):

serve **qualcuno che abbia un'idea di quali sono i dati** a disposizione; e **qualcuno che possa facilmente relazionarsi** con i vari dirigenti responsabili di ogni dataset, eventualmente anche per capire il processo grazie al quale i dati sono in possesso della pubblica amministrazione; serve **qualcuno che abbia competenze legali**, in particolare di proprietà intellettuale e – eventualmente in misura minore – di tutela dei dati personali (e magari di diritto amministrativo, ma questo aspetto può spesso essere coperto da altri membri della squadra); è inoltre utile che **qualcuno abbia un'idea di come i dati sono gestiti tecnicamente** e che sappia dare qualche indicazione di base sui modi tecnici per mettere i dati a disposizione, in maniera che siano facilmente riutilizzabili; possibilmente, **qualcuno dovrebbe conoscere (o studiare) gli aspetti principali del movimento open data**, anche a livello internazionale, per non scordarsi mai i principi chiave e per ricordare agli altri le opportunità dietro l'angolo.

Fate cose semplici

Soprattutto, partite da progetti piccoli e semplici. Non è necessario aprire tutti i dati; non è necessario aver mappato tutti i dataset che si possiedono prima di partire; non è necessario "aprire" prima i dati più interessanti, se farlo solleva problemi particolarmente complessi. E, sì, si può partire dai dati nella sfera di competenza e controllo dei politici e dei dirigenti più aperti, innovativi o, semplicemente, cui siete riusciti a spiegare meglio a cosa serva aprire i dati: quasi tutti i progetti di successo che conosciamo, in effetti, hanno seguito questa strada.

Naturalmente, siate onesti: i potenziali riutilizzatori si renderanno conto se, tra tutti i dati, starete mettendo a disposizione solo quelli che non interessano a nessuno dentro la vostra amministrazione o non rischiano neanche lontanamente di toccare temi delicati dal punto di vista politico. **Siate trasparenti e chiarite che quello che state facendo è solo l'inizio di un processo complesso: l'esperienza suggerisce che il fatto di aver cominciato, anche se a piccoli passi, sarà apprezzato**. Se farete annunci roboanti, invece, qualcuno potrebbe far giustamente osservare che "il Re è nudo!" e che il vostro progetto open data sembra più che altro un'operazione di marketing.

Per partire da cose semplici, infine, non è necessario abbandonare progetti ambiziosi: preferite, però, **progetti modulari, i cui primi moduli offrano già qualche concreto risultato** in termini di dati disponibili al pubblico.

Date aggiornamenti frequenti

Una delle ragioni per fare cose semplici è poterle fare abbastanza in fretta. Un'altra è apprendere dall'esperienza. Rilasciando spesso quello che avete a disposizione avrete due grandi vantaggi: mantenere altro l'interesse e rendere credibile la vostra strategia modulare (poiché ci saranno frequentemente nuovi piccoli dataset a disposizione), facendovi perdonare anche da chi non ha ancora avuto la base di dati che gli interessava; e imparare dai vostri errori o dalle esperienze positive. I dati aperti sono un campo nuovo, sia per chi li mette a disposizione che per chi li riusa: imparare dall'interazione reciproca presenta maggiori probabilità di successo di molti piani di lungo periodo.

Coinvolgete le persone. Fatelo fin da subito, fatelo spesso

Parlate con i riutilizzatori attuali e potenziali. L'idea non è quella di raccogliere tutti gli *stakeholder* potenzialmente interessati attorno ad un tavolo di lavoro. In astratto, questo potrebbe servire, ma i tavoli di lavoro sono spesso percepiti – forse non del tutto a torto – come sostitutivi del lavoro stesso... L'invito che facciamo qui è semplicemente a **confrontarsi con le esigenze di chi i dati li potrà poi riutilizzare**. Nel farlo, ricordate che alcuni riutilizzatori sono "indiretti", ovvero degli intermediari dell'informazione: solo

pochi cittadini e imprese saranno in grado di riutilizzare direttamente i dati pubblici. Molti altri, potranno accedere a questi dati grazie al lavoro di questi pionieri capaci di riorganizzare, trasformare e presentare nuovamente i dati. L'intermediazione e a volte l'aggregazione e la scelta di punti di vista parziali sui dati stessi, dunque, restano necessari: il vantaggio dell'approccio open data è la presenza (almeno potenziale) di una pluralità di intermediari e di punti di vista.

C'è chi può aiutare

Nell'individuare i propri interlocutori, ci sono alcuni soggetti da tenere in considerazione, ad esempio: i rappresentanti di associazioni e comunità come Wikimedia Italia ("quelli di Wikipedia"), OpenStreetMap o GFOSS (l'Associazione italiana per l'informazione geografica libera). O, ancora, i vari esponenti della community open data italiana, dall'Associazione Italiana per l'Open Government a SpaghettiOpenData passando per Linked Open Data Italia.

Il Centro NEXA su Internet e Società del Politecnico di Torino, oltre che per le sue attività accademiche, è attivo in questo campo come ente affiliato a Creative Commons internazionale (si veda anche la sezione 2.3). Open Knowledge Foundation Italia è un altro interlocutore potenziale, che tra l'altro gestisce, insieme con il Centro NEXA, CKAN Italia, il primo portale open data di community in Italia.

A livello accademico, per il momento, i centri di competenza maggiori si trovano probabilmente in Piemonte (dove il progetto EVPSI coinvolge l'Università di Torino, il Centro NEXA, la Fondazione Rosselli e l'Università del Piemonte Orientale, mentre il Centro NEXA coordina LAPSI, il network tematico europeo sugli ostacoli giuridici al riutilizzo dell'informazione del settore pubblico), in Toscana, dove la Scuola Superiore Sant'Anna e l'ITTIG del CNR hanno portato avanti ricerche e seminari sul tema e in Lombardia, dove sono attivi gruppi di lavoro su questi temi alle Università Bocconi e Bicocca. A livello di pubbliche amministrazioni, meritano specifica menzione il tavolo di lavoro regionale piemontese (formato da Regione Piemonte, CSI, Top-IX, Centro NEXA, CSP e Istituto Superiore Mario Boella) e i gruppi di lavoro che hanno dato vita ai portali open data dell'Emilia Romagna, del Veneto, della Provincia di Trento, della Lombardia e del Comune di Firenze, quest'ultimo anche in collaborazione con l'associazione Wikitalia.

Non chiedetevi "a chi può servire o interessare?"

Spesso ci si chiede come possano i cittadini usare in qualunque modo sensato dei database di enorme ricchezza e complessità. Ancora una volta, non sottovalutate gli intermediari della Rete: può darsi che siano in grado di creare interfacce tramite le quali molti altri cittadini (o direttamente i loro smartphone e computer!) saranno in grado di fruire dei vostri dati, eventualmente arricchendoli e correggendoli. Inoltre, lo ripetiamo: se non ci sono buone ragioni per non condividere un dato e non sapete a cosa possa essere utile quel dato, allora c'è un'ottima ragione per "aprirlo": sicuramente, se qualcuno lo riuserà, si tratterà di un riutilizzo a cui voi difficilmente avreste potuto pensare!

Non spaventatevi troppo e, se lo siete, parlatene

Capita spesso che qualcuno sollevi problemi apparentemente difficili o impossibili da risolvere. Quasi sempre,

questi problemi hanno una soluzione, anche se capita che tale soluzione passi per una riduzione della qualità informativa dei dati (anonimizzazione, aggregazione – inclusa la pubblicazione di medie o dati statistici – o simili procedure che fanno allontanare dal dato grezzo). In caso di incertezza, è utile confrontarsi con chi ha esperienza di progetti relativi ai dati aperti, perché è probabile che i problemi di ciascuno siano già stati affrontati (e superati) da altri. Senza nulla togliere alla competenza degli esperti (ad esempio di problemi legali) che stanno dentro ciascuna istituzione, spesso è interessante e produttivo mettere queste persone in contatto con chi sta fuori dall'istituzione stessa e ha esperienza di open data, perché il punto di vista da cui affrontare le questioni relative ai dati aperti è molto diverso da quello tradizionalmente adottato per la gestione interna della PA.

Non siate integralisti, in nessun senso

I dati pubblici sono tanti e diversi, è dunque naturale che non tutti possano essere trattati allo stesso modo. Quando introducete qualcuno a questo tema, potete certamente andare dritti ai principi generali relativi all'apertura e alla libertà massime. Tuttavia, ricordate e spiegate sempre che non ci sono questioni di principio, ma solo questioni di buon senso; che tutte le regole hanno eccezioni, ma che fare eccezioni non dovrebbe essere la regola.

Nell'incertezza, è quasi sempre possibile e sensato adottare più di una soluzione: offrire il dato in più formati, applicare più licenze (ciascuna delle quali non esclusiva). L'unica cautela da adottare è evitare la confusione: un certo formato e licenza (possibilmente i più standard, aperti e liberi) dovrebbero essere quelli più chiaramente pubblicizzati, offrendo comunque le alternative per massimizzare la libertà dei riutilizzatori.

Rendete i dati disponibili gratuitamente

Di fatto, questo principio è una sorta di corollario dell'uso di licenze libere, poiché un dato **liberamente riutilizzabile** e redistribuibile non è facile da monetizzare (ossia dare in licenza ad un prezzo che renda conveniente la transazione). Tuttavia, ci sono altre ragioni per distribuire gratuitamente i dati. La letteratura economica su questo tema, in effetti, ha confermato il valore della regola aurea "Prezzo = Costo Marginale" in questo campo. E, data la distribuzione online, il costo marginale è sostanzialmente zero. Infine, e nella pratica questa potrebbe essere l'argomentazione più persuasiva, non è raro che il costo di esigere una tariffa sia maggiore del ricavo generato dalla stessa!

Quando farsi pagare per il riutilizzo dei dati non conviene

Il Sistema Informativo Territoriale della Regione Piemonte comprende una serie articolata di repertori cartografici, aggiornati nel tempo. Si tratta di un insieme di dati geografici, rappresentati nella forma di mappe che riportano informazioni geometriche, topologiche e descrittive del territorio regionale. Questi dati – che rappresentano un'istantanea del territorio regionale e degli oggetti, naturali e artificiali, che lo compongono – si prestano a svariate forme di riutilizzo di particolare importanza, dalla creazione di nuove mappe (con l'aggiunta di ulteriori informazioni) alla georeferenziazione dei servizi digitali. L'aggiornamento dei dati cartografici, pur facendo parte del mandato pubblico, comporta per l'amministrazione regionale un ingente costo. Per dare una misura concreta, la realizzazione della carta di un grande centro urbano (a scala 1:2000, con un aggiornamento tendenzialmente necessario ogni due anni) comporta una spesa di circa 50 euro

all'ettaro.

Prima dell'approvazione delle *Linee Guida relative al riutilizzo e all'interscambio del patrimonio informativo regionale* (novembre 2010), le quali prevedono in Piemonte un rilascio libero e gratuito dei dati dell'amministrazione pubblica, parte dei repertori geografici veniva rilasciata a titolo oneroso, secondo uno specifico tariffario e, naturalmente, condizioni di rilascio compatibili con quanto prescritto dalla direttiva europea INSPIRE (relativa appunto ai dati geografici). I ricavi netti annui stimati erano compresi tra i 10mila e i 20mila euro. I costi cagionati dalla natura onerosa e manuale delle transazioni (prive di un canale di distribuzione digitale) – ossia le risorse umane dedicate, nonché le spese di fatturazione e altre spese accessorie – assommavano invece a circa 50mila euro all'anno, rendendo dunque, di fatto, antieconomica l'offerta di dati a titolo oneroso e incentivando un rilascio libero attraverso la rete.

La scarsità di risorse non consente peraltro di effettuare gli aggiornamenti dei database geografici con la frequenza che sarebbe desiderabile: per questo, meccanismi virtuosi di collaborazione con i riutilizzatori delle mappe potrebbero rappresentare un'opportunità molto più interessante della "vendita" del dato stesso.

2.2 I passi chiave nel processo di apertura

Ecco i principali passi per aprire una base di dati (alcuni dei quali saranno poi approfonditi in appositi capitoli):

1. Individuare uno specifico dataset:

analizzare i punti successivi in astratto è difficile o impossibile, per cui è utile discutere sempre di dati abbastanza ben determinati, di cui sia possibile mostrare un campione ai vari possibili interlocutori (giuristi, dirigenti, amministratori, potenziali riutilizzatori);

Sempre un passo alla volta!

È importante (poter) discutere di dati reali, visualizzarli, parlare con le persone responsabili della loro raccolta, ecc. È quindi indispensabile individuare alcuni dati. Non cadete tuttavia nell'errore di voler mappare tutti i dati disponibili prima di cominciare ad aprirne qualcuno: salvo che siate così fortunati da avere già un catalogo ben organizzato, l'opera rischierebbe di essere faraonica e vi impedirebbe di conseguire risultati concreti nel breve periodo, così facendo perdere entusiasmo a tutti.

2. Individuare quali diritti esistano sui dati:

da dove vengono i dati? Chi li ha raccolti? Con quali risorse? Ci sono (stati) contratti di qualche tipo che regola(va)no l'attività di raccolta e/o l'uso dei dati stessi? Parte di questi dati sono personali? O addirittura sensibili? Ci possono essere diritti di proprietà (ad esempio informazioni commerciali segrete)?

Se troppe di queste domande sembrano difficili o impossibili da rispondere, anche con l'aiuto dell'ufficio legale del vostro ente, prendete in considerazione l'idea di scegliere un altro dataset, in particolare se siete all'inizio della vostra attività di apertura dei dati.

3. Applicare una licenza aperta:

quali diritti volete concedere ai riutilizzatori? A che condizioni? Chi è titolato a rilasciare la licenza in questione? Abbiamo già spiegato perché la scelta della massima apertura sia normalmente ben difendibile (si veda il paragrafo 3.1). In generale, è opportuno che esista un documento, condiviso e formalmente approvato, che legittimi la procedura di licenza dei dati. Può essere uno dei passaggi istituzionali più utili nell'evoluzione da sporadici progetti pilota open data ad una vera politica dei dati aperti;

Ma chi licenzia i dati pubblici?

Riguardo la procedura necessaria alla concessione di una licenza, è possibile che – nelle prime fasi di un progetto open data – trovare una risposta condivisa non sia banale. Alcuni enti riterranno che un qualche tipo di determina dirigenziale o ordine di servizio sia più che sufficiente (in particolare nel caso di progetti pilota su specifici insiemi di dati), altri vorranno portare l'iniziativa a conoscenza dei massimi organismi rappresentativi che governano l'ente stesso (il che è probabilmente una buona idea nel lungo periodo, al fine di dare la massima legittimità e stabilità all'iniziativa open data, dandole anche una specifica valenza democratica). Una via di mezzo, rappresentata da **una delibera di un organo di governo, quale la giunta comunale o regionale, può essere uno strumento operativamente efficace per far partire la prima fase di un progetto strutturato** (e ci sono molti precedenti in questo senso).

4. Mettere i dati a disposizione online:

per cominciare, un semplice archivio compresso (per i file di una certa dimensione), in cui i dati siano disponibili, in blocco e in un formato il più possibile diffuso, va benissimo.

Questo passo è normalmente banale. Infatti, quasi solo i progetti di grande successo possono permettersi il lusso di porsi problemi relativi ad un eccessivo carico dei loro server causato dal download dei dati. Inoltre, qualora succeda, poter dire che l'eccessiva domanda ha messo in ginocchio i propri server è spesso una pubblicità più positiva che negativa (a condizione di non utilizzare per la distribuzione server che ospitino servizi critici o importanti per i cittadini). Se pensate che l'eccessiva domanda sia comunque una criticità da prevedere, prendete in considerazione strumenti distribuiti[5];

Download in blocco vs. API

Il modo più banale, ma anche uno dei più efficaci, per mettere a disposizione dei dati nell'era di Internet, è la semplice pubblicazione online, su un sito Web, di un file statico (eventualmente sotto forma di archivio compresso), il cui link sia chiaramente disponibile secondo lo standard HTTP.

Un altro modo è rappresentato da una *Application Programming Interface* (API), che renda le informazioni disponibili a richiesta, in maniera più o meno elaborata (ad esempio, permettendo di scaricare solo un certo sottoinsieme dei dati, che interessano all'utente in quel momento, o solo gli aggiornamenti degli ultimi 5 minuti, ecc.). Rendere i dati disponibili tramite un'API, laddove possibile, è utile poiché può

5 Il riferimento è a strumenti di tipo *peer-to-peer*: le reti di *file sharing* vengono usate da tempo per distribuire *software* libero e altri contenuti liberamente riutilizzabili. Utilizzarle per distribuire grosse moli di dati è una soluzione semplice e a basso costo, col pregio di essere scalabile, perché diventa sempre più efficiente quanti più soggetti decidono di scaricare il dato.

facilitare la vita a chi voglia realizzare delle *apps* o *mashups* di vario tipo (specie se girano su un telefono cellulare).

Malgrado i possibili vantaggi di un'API, il rilascio dei dati (anche) in blocco, in modo che siano facili da scaricare con un semplice link è sempre fondamentale, poiché assicura che tutti possano facilmente ottenere una copia dei dati e redistribuirla; rende facile ad altri sviluppare i propri servizi (ed eventualmente le loro API specifiche, oppure rappresentazioni degli stessi dati in altro formato, ecc.).

5. Far trovare i dati:

non nascondete i vostri dati: possono avere grandi potenzialità, ma se nessuno sa che esistono e dove sono, resteranno inutilizzati. Spendete dunque qualche energia anche per rendere i vostri dati facilmente individuabili sul Web. Volendo, potete creare un vostro apposito sito/catalogo o portale (a volte detto anche Information Asset Register). Oppure potete verificare se ne esiste uno, ad esempio, nella vostra regione (dati.piemonte.it è un caso di portale regionale aperto a varie pubbliche amministrazioni e istituzioni piemontesi). Per il momento, non esistono in Italia portali/cataloghi pubblici nazionali dedicati ai dati aperti, esiste tuttavia un portale gestito da una comunità di utenti: IT.CKAN.NET (sezione italiana dell'internazionale CKAN.NET). Chiunque può inserire su CKAN le informazioni relative a database aperti, rendendoli così più facili da scorprire. Inoltre, gli amministratori di CKAN saranno probabilmente molto ben disposti a mettere a disposizione della vostra istituzione un apposito gruppo, che vi permetterà di mantenere il controllo sulla versione ufficiale delle informazioni relative ai dataset che distribuite.

Serendipità

L'apertura dei dati, è una scommessa sulla serendipità, sulla possibilità che un'informazione e chi sa come usarla possano incontrarsi, magari senza che il secondo sappia chiaramente di essere alla ricerca della prima. Far sì che i dati siano collegati in molti modi diversi, indicizzati dai motori di ricerca (generalisti e specializzati sui dati pubblici) è un modo per aumentare le chance di questo incontro fortuito!

I like!

In caso abbiate a disposizione un catalogo (anche incompleto) dei vostri dataset, è utile metterlo a disposizione, eventualmente fornendo degli strumenti per esprimere interesse verso una certa base di dati (anche non ancora disponibile per l'accesso o il riutilizzo). Tale interesse può essere espresso in vari modi, da un semplice bottone "Mi interessa!" (o addirittura l'ormai celebre "I like!" di Facebook) a più articolati meccanismi di voto o form di contatto, in cui spiegare anche le ragioni dell'interesse. (Ricordate solo che tali ragioni, di norma, sono interessanti per stabilire la priorità nella messa a disposizione, ma non dovrebbero essere richieste obbligatoriamente o vincolare i tipi di riutilizzo ammissibili.)

2.3 Licenziare i dati: pillole di diritto della proprietà intellettuale

Nella maggior parte dei casi, anche se nessuno ha mai esplicitamente reclamato alcun diritto su una banca di dati, è ragionevole presumere che esistano alcuni diritti di proprietà intellettuale (e *in primis* il diritto d'autore e il diritto *sui generis*)[6] e che sia necessario chiedere il permesso del titolare dei diritti per compiere la maggior parte delle operazioni di riutilizzo che possano venire in mente (estrazione di dati, riproduzione, aggiornamento, creazione di nuove banche dati che ne combinino di vecchie, etc.). Di regola, il fatto che tali atti siano compiuti per fini non-commerciali non legittima l'utilizzazione (il che vale in generale per tutti i materiali protetti diritto d'autore e dunque anche nel nostro caso). In altre parole, il modello in cui ricadono automaticamente le banche di dati è quello "Tutti i diritti riservati".

È dunque necessario porsi due questioni: 1) può l'ente pubblico disporre legalmente dei dati in questione, usandoli liberamente ed eventualmente anche ri-licenziandoli come ritiene opportuno? 2) **Se sì, com'è opportuno che l'ente si comportarti nel gestire i propri diritti?**

La prima domanda, che riguarda la cosiddetta *rights clearance*, rientra tra le questioni tecnico-giuridiche che non trattiamo in dettaglio in questo lavoro (è anche per questa ragione che è importante avere un giurista nella squadra operativa dietro ogni progetto open data). Per ora, ipotizzeremo che la risposta sia semplicemente sì, il che accade, per esempio, quando una banca dati è creata da dipendenti di un ente pubblico, con fondi dell'ente stesso.

Rights clearance

Per analizzare lo status giuridico di un database, sarà necessario rispondere a quesiti del tipo: Da dove vengono i dati? Chi li ha raccolti? Con quali risorse? Ci sono (stati) contratti di qualche tipo che regola(va)no l'attività di raccolta e/o l'uso dei dati stessi? (Inoltre, sarà opportuno chiedersi: Parte di questi dati sono personali? O addirittura sensibili? Ci possono essere segreti aziendali od altri diritti di proprietà intellettuale di terzi?)

Le risposte alle domande di cui sopra saranno utili ad un giurista per stabilire se e come i dati possano essere aperti" Tuttavia, ricordate che, **specie all'inizio di un progetto open data, potrebbe essere saggio scegliere un dataset meno "ostico", qualora l'attività di *rights clearance* risultasse troppo complessa**, malgrado l'assistenza dell'ufficio legale del vostro ente.

Come abbiamo accennato, a causa della nostra legge sulla proprietà intellettuale, lo scenario base è del tipo "tutti i diritti riservati". In altre parole, per chi conosce il diritto, è un po' **come se, pubblicando un dato, voi scriveste automaticamente anche "con questi dati non si può far nulla (salvo guardarli, se li stai trovando online – ma occhio che non ti ho detto che puoi salvarteli sul disco fisso!)"**. Ovviamente, qualcuno potrebbe decidere di usare comunque i dati (anche perché potrebbe non sapere che è probabilmente illecito farlo); ma se l'obiettivo di un ente pubblico è aumentare in modo significativo l'uso delle risorse informative che possiede,

6 e banche di dati possono essere tutelate dal diritto d'autore e/o da un particolare diritto – detto, appunto, diritto sui generis sulle banche di dati. Senza scendere in eccessivi dettagli, le banche dati in qualche modo "creative" sono tutelate dal diritto d'autore, mentre sia le banche dati creative che quelle meramente e meccanicamente compilative possono essere tutelate dal diritto sui generis, qualora la loro costituzione abbia comportato un investimento rilevante.

è sempre bene associare ai dati un'esplicita licenza (o almeno una liberatoria, che espliciti la facoltà per chiunque di riutilizzare i dati liberamente). **L'incertezza è nemica del riutilizzo**. Inoltre, anche se un detentore di diritti può concedere (o negare) licenze caso-per-caso, eventualmente decidendo di realizzare una licenza *ad hoc* per ogni riutilizzatore, **concedere molte autorizzazioni caso-per-caso è una pessima idea**, perché richiederebbe risorse da parte del detentore dei diritti, oltre a scoraggiare i riutilizzatori.

Notate che anche la presenza di termini per il riutilizzo poco chiari potrebbe avere effetti scoraggianti. Se, per capire cosa si possa davvero fare, serve un avvocato, probabilmente nessuno riutilizzerà i vostri dati. È dunque utile adottare **strumenti per ridurre i costi di transazione**, ovvero i costi monetari, ma anche le perdite di tempo e l'incertezza, sia per i detentori dei dati, sia per i riutilizzatori. Questi strumenti sono le **licenze standard di diritto d'autore (e connessi)**.

Una licenza è un documento che descrive quali, tra le attività normalmente riservate al detentore dei diritti (il licenziante), possano essere compiute da chi riceve i dati sotto i termini della licenza stessa (il licenziatario). Eventualmente, la licenza descrive alcune condizioni per godere dei diritti concessi. Come abbiamo già discusso, definendo a che condizioni i dati siano davvero "aperti", la scelta ottimale è tra "nessuna condizione" e "a condizione che si riconosca l'attribuzione, ovvero si menzioni l'origine dei dati". Combinando l'opportunità di utilizzare gli strumenti più standard e diffusi con le condizioni di cui sopra, è lecito raccomandare di restringere la scelta delle licenze al seguente elenco: la licenza/liberatoria **Creative Commons Zero (CCo)**, per chi voglia scegliere l'approccio "nessuna condizione"; la licenza **Creative Commons Attribuzione (CC BY)** per chi sia voglia richiedere l'attribuzione.

Adozione di licenze aperte come standard per la PSI

Sono ormai numerosi i paesi che hanno adottato – come unica soluzione per le loro informazioni pubbliche o, quantomeno, come soluzione standard – una licenza aperta.

Negli **USA, le informazioni e i contenuti generati del settore pubblico (a livello federale) ricadono automaticamente nel pubblico dominio** e sono dunque liberamente riutilizzabili senza alcun vincolo e per legge. Questa è chiaramente la soluzione aperta per eccellenza, perché non richiede scelte ulteriori e si applica automaticamente a tutta la PSI.

Nel **Regno Unito**, è stata sviluppata una licenza ad hoc per l'informazione del settore pubblico, **la Open Government Licence. Si tratta di una licenza che richiede la sola citazione dell'origine dei dati** e che è esplicitamente compatibile con le licenze CC BY ed Open Data Commons Attribution (ODC BY). Con una scelta simile, in **Nuova Zelanda, la licenza Creative Commons Attribuzione (CC BY) è stata scelta come soluzione di *default***. In pratica – in assenza di buone ragioni per fare altrimenti – la scelta ricade automaticamente su quella licenza.

In Italia, **Regione Piemonte ha scelto CCo come propria licenza di *default***. Ad oggi, l'unica alternativa praticata è stata CC BY, prescelta per alcuni dati cartografici.

Il requisito di attribuzione non richiede particolari commenti e consiste nell'obbligo di citazione dell'origine/fonte dei dati (ovviamente, con modalità che non traggano in inganno l'utilizzatore – ad esempio, un dataset modificato

dovrà sempre essere identificato come tale e mai spacciato per l'originale). Al contrario, **il requisito di condivisione allo stesso modo è un tema complesso e delicato**. Questa clausola (detta anche *share-alike*, *copyleft* o virale) consiste nell'obbligare il riutilizzatore a licenziare qualsiasi lavoro realizzato a partire dall'originale con la medesima licenza adottata per l'originale (sicché chiunque altro resta a sua volta libero di apportare ulteriori modifiche e la catena di condivisione si prolunga, da cui il nome clausola virale). Anticipiamo fin d'ora che – per quel che riguarda i dati delle pubbliche amministrazioni – **non tutta la comunità open data è d'accordo sull'opportunità di chiamare aperti dati che siano vincolati da una clausola share-alike**.

Usare con cautela: Condividi-allo-stesso-modo

Spesso, i dati generati dalle *community online* sono rilasciati sotto licenze del tipo condividi-allo-stesso-modo. Questa è una **scelta di auto-difesa della comunità**, contro soggetti esterni che tentino di appropriarsi del lavoro della comunità stessa, tramite strategie che passino per la creazione di una versione modificata (e migliorata) del lavoro della community. Per una comunità di utenti, che si unisce intorno all'idea di un progetto comune che rappresenta anche l'obiettivo attorno al quale si aggrega la comunità stessa, questa scelta può essere appropriata. È **più dubbio se lo stesso valga per un'amministrazione pubblica**, che si ritrova a possedere dati generati come "effetto collaterale" di un suo compito istituzionale. In particolare, mentre lo scopo principale di una *community online* è far progredire il progetto attorno al quale si è costituita, **i dati pubblici sono messi a disposizione per rendere possibili riusi inattesi, spesso in combinazione con altri insiemi di dati**. Per questo motivo, l'inserimento della clausola di condivisione-nello-stesso-modo è un'idea criticabile. Ogni licenza *share-alike*, infatti, è compatibile con se stessa e tendenzialmente incompatibile con altre licenze *share-alike* (ad esempio, CC BY-SA e ODC ODbL sono mutualmente incompatibili). **Il risultato rischia di essere una sorta di "Babele delle licenze", in cui numerosi insiemi di dati non possono essere uniti e mescolati**.

Se comunque si decidesse di adottare una licenza di tipo "condividi-allo-stesso-modo", CC BY-SA avrebbe il chiaro vantaggio di essere lo standard dominante (ad esempio, perché utilizzata da Wikipedia e progetti collegati, come dbPedia). Tra le altre principali licenze contenenti una clausola condividi-allo-stesso-modo, la ODC ODbL (*vedi box successivo*) presenterebbe il vantaggio di essere più robusta in termini di possibilità di far valere con successo la clausola condividi-allo-stesso-modo in alcune situazioni (poiché è stata scritta appositamente per le banche di dati, mentre le licenze CC sono più generiche). La IODL versione 1.0 (*vedi box successivo*) presenterebbe invece il vantaggio di contenere un'esplicita clausola di compatibilità, che permette di mescolare dati ottenuti sotto la IODL con dati licenziati con una licenza compatibile (CC BY-SA o ODC ODbL), ri-licenziandoli con la seconda licenza. Forse, se proprio si vuole usare una licenza "virale", la sua clausola di compatibilità rende la IODL 1.0 la scelta preferibile (almeno in Italia).

E Open Data Commons? E la Italian Open Data Licence 2.0?

I lettori con maggior familiarità con le licenze aperte si staranno ponendo le domande di cui sopra (e altre, troppo tecniche per essere discusse in questa sede...). Per rispondere loro è giusto chiarire subito che esistono due alternative a CC0 e CC BY, realizzate da Open Data Commons (http://opendatacommons.org/) che risultano tecnicamente del tutto adeguate (e storicamente anche più "vicine" al movimento open data). La ragione per cui raccomandiamo gli strumenti realizzati da Creative Commons è semplicemente che questi sono più diffusi, meglio documentati sul Web, e forse leggermente più comprensibili, oltre che – e questo non è un aspetto da trascurare – tradotti in italiano[7].

Per quanto riguarda la versione 2.0 della Italian Open Data Licence sviluppata da Formez PA (http://www.dati.gov.it/iodl/2.0/), sebbene presenti la sola clausola di attribuzione, noi non la raccomandiamo esplicitamente, per ragioni di opportunità legate ad aspetti di standardizzazione tra licenze, anche in relazione al contesto internazionale. Così come non raccomandiamo le licenze di tipo condividi-allo-stesso-modo (o *share-alike*) sviluppate da Open Data Commons (Open Database License, ODC ODbL) o da Creative Commons (Attribuzione Condividi-allo-stesso-modo, CC BY-SA), per le ragioni, legate all'interoperabilità tra licenze, descritte nel precedente box.

Proprio rispetto alla necessità di interoperabilità tra licenze, vanno citate iniziative di communities che spingono verso l'adozione di una singola licenza per i dati pubblici aperti in Europa (si veda, ad esempio, la petizione promossa dalla comunità Open Data spagnola – http://chn.ge/1LUJ42). Questo approccio pare compatibile con i principi espressi nell'Agenda Digitale europea, specie in relazione alla creazione di un singolo mercato digitale. Nel prossimo futuro, è dunque possibile che i decisori europei intraprendano il percorso verso un unico modello standard di licenza.

Per applicare una licenza ad una base di dati distribuita online, la prassi è che basti indicare la licenza stessa nella pagina dalla quale si effettua il download[8]. Inoltre, esistono opportuni accorgimenti tecnici, per far sì che l'associazione tra file e licenza sia chiara anche alle macchine (p. es. ai motori di ricerca)[9].

Le licenze sono flessibili. (Ma occhio a non renderle contorte!)

Tra le possibilità offerte dal diritto d'autore, c'è anche quella di utilizzare **licenze non esclusive** (in effetti, tutte le licenze standard da noi menzionate sono di questo tipo). In tal caso, **il detentore**

7 Per completezza va anche menzionato un vantaggio della soluzione Open Data Commons Attribution (ODC BY), che è rappresentato dal fatto che questa licenza fa leva anche sul diritto sui generis e non sul diritto d'autore per pretendere il rispetto del requisito dell'attribuzione.

8 Si noti che le licenze relative a diritti d'autore devono essere provate per iscritto. Tuttavia, l'assenza di licenza equivale allo status "tutti i diritti riservati", per cui un licenziatario che violi i termini della licenza indicata sul sito difficilmente potrà giovarsi del fatto che la licenza vada provata per iscritto. Questo aspetto, dunque, può essere ignorato, almeno in prima approssimazione.

9 Per esempio, utilizzando il "license chooser" (selettore di licenze) di Creative Commons (http://creativecommons.org/choose/?lang=it) si riceve automaticamente un brano di codice da inserire nelle proprie pagine. Per maggiori dettagli: http://wiki.creativecommons.org/CC_REL.

dei diritti può offrire diverse licenze allo stesso tempo, ad esempio garantendo una licenza standard a chiunque sia interessato a rispettarne i termini, ma affiancandovi la possibilità di accordi speciali in casi particolari. Ovviamente, a seconda dei termini di ciascuna delle licenze scelte, questa possibilità può o meno avere senso. Per esempio, un ente cartografico potrebbe offrire gratuitamente una licenza di tipo condividi-allo-stesso-modo per le proprie mappe, ma potrebbe anche offrire, dietro compenso, un'altra licenza che non includa tali limitazioni. Questa seconda licenza, per esempio, potrebbe essere scelta da operatori commerciali, che vogliano mescolare propri dati geografici con quelli dell'ente cartografico, utilizzando poi una licenza proprietaria per il risultato finale. (Notate che questa non è necessariamente una soluzione raccomandabile: è solo un esempio di combinazione possibile di licenze).

Un'altra possibilità – forse più raccomandabile della precedente – passa per l'**uso di licenze a pagamento, che includano servizi aggiuntivi (un *call center*, la possibilità di download in larga banda, una garanzia o certificazione aggiuntiva di qualche tipo sui dati)**. Qui, il caveat è contro stratagemmi che deteriorino la qualità dell'offerta di base rispetto ai principi discussi in questo libro bianco, al fine di creare articifialmente domanda per la versione premium".

Il principio cardine è che i dati, grezzi ed aggiornati, siano disponibili a tutti, gratuitamente, sotto una licenza aperta. A questa condizione, può essere sensato valutare ulteriori licenze per casi particolari.

...e se qualcosa non è protetto da alcun diritto?

Per semplicità, abbiamo ipotizzato che tutti i dati siano protetti dalla proprietà intellettuale, ma ciò non è vero. Per fortuna, alcuni database sono liberi e riutilizzabili da parte di chiunque. Se vi trovate in questo caso e avete creato voi il database, potete scegliere di **applicare comunque la liberatoria CCo** (che, essendo del tipo "nessun diritto riservato", **non rischia di creare confusione e anzi fuga qualsiasi dubbio**). Se, invece, i dati provengono da altri, lo strumento per etichettarli come liberi (senza licenziarli, perché non ne avreste la facoltà) è il **Creative Commons Public Domain Mark** (http://craetivecommons.org/publicdomain/mark/1.0). Questa soluzione, per esempio, è proposta anche dalle guidelines governative neozelandesi (si veda www.e.govt.nz/policy/nzgoal).

2.4 Rendere i dati disponibili: pillole di tecnologia

I dati aperti hanno bisogno di essere tecnicamente riutilizzabili, oltre che legalmente tali.

Abbiamo già detto dell'importanza di rendere possibile un download, possibilmente gratuito, via Internet. E abbiamo sottolineato quanto sia importante che i dati possano essere scaricati in blocco, ad esempio come un unico archivio compresso. A scanso di equivoci, rendere i dati disponibili solo tramite servizi web, ad esempio in pagine con menù a discesa, che restituiscono alcuni dati, sulla base delle scelte dell'utente, non è una modalità

adeguata a facilitare il riutilizzo (anche se il servizio può essere utilissimo in quanto tale).

Come già immaginerete è anche opportuno che i formati dei file siano aperti.

Formati aperti, secondo la legge italiana

> "Le informazioni sui formati aperti sono consultabili all'indirizzo: www.digitpa.gov.it/formati-aperti. Secondo le Linee guida per i siti web della PA 2010, previste ai sensi della Direttiva 8/2009 del Ministro per la pubblica amministrazione e l'innovazione, è raccomandato l'uso dei seguenti formati aperti e standardizzati: HTML/XHTML per la pubblicazione di informazioni pubbliche su internet; PDF con marcatura (secondo standard ISO/IEC 32000-1:2008); XML per la realizzazione di database di pubblico accesso ai dati; ODF e OOXML per documenti di testo; PNG per le immagini; OGG per file audio; Theora per file video"[10].

I formati aperti sono una condizione necessaria (e comunque assai raccomandabile), ma non sempre sufficiente. La **condizione sufficiente è l'uso di un formato aperto & machine-readable (leggibile in modo automatico)**. E la possibilità di processare facilmente in automatico i file è importante, perché a volte facilita il riutilizzo anche più di un formato aperto. Per esempio, una tabella di Microsoft Excel (formato XLS) non è aperta, ma è *machine readable* e, a fini pratici, risulta infinitamente più utile di un PDF, che è un formato aperto, ma dal quale può essere molto difficile estrarre dati ben organizzati!

10 Fonte: http://www.forumcad.ancitel.it/Quesito.cfm?id=279.

Alcuni esempi di ottimi formati machine readable (ed aperti) sono l'XML e il CSV (Comma Separated Value).

Occhio al PDF!

> A rigore, il formato PDF è aperto, nel senso che si tratta di uno standard ben documentato ed utilizzabile da parte di chiunque. Questo formato va dunque benissimo per distribuire documenti, che vadano letti così come sono stati scritti (diciamo che va bene per l'accesso). Ma è pessimo per estrarre dati strutturati. Per chi è interessato al riutilizzo, ricevere una tabella in formato PDF è quasi equivalente ad averla solo su carta (perché in entrambi i casi si rischia di doversela riscrivere a mano in un foglio di calcolo o database opportuno).

In caso di incertezza, la parola d'ordine resta "DATI GREZZI, ORA!". Buttate fuori quel che c'è, nel formato che anche a voi è comodo utilizzare (eventualmente anche l'XLS o il DOC). Inoltre, pubblicate possibilmente una conversione in un formato aperto, la quale può spesso essere realizzata automaticamente e a costi sostanzialmente nulli. (Inoltre, mentre è importante non pubblicare dati che legalmente non si sarebbero potuti pubblicare, sbagliare il formato o lo standard non è un problema altrettanto grave: basta segnalare che si tratta di un esperimento, di una prova, di una release alpha... insomma, basta chiarire che in futuro la tecnica di rilascio dei dati potrebbe cambiare, per essere più aperta e più standard.)

Metadati: i dati sui dati

Quando pubblicate un dato, considerate che è spesso necessario spiegare il significato delle colonne e dei campi delle vostre tabelle a terzi che non conosco nulla dei vostri sistemi informativi interni (per esempio, se ci sono codici è necessaria una tabella apposita, che funga da riferimento per espandere i codici stessi). Ovviamente, questo è un aspetto che sarà ben chiaro a chi gestisce i vostri database. Tra le informazioni che può essere utile fornire ci sono quelle sulla qualità dei dati (qual è il livello di precisione, la frequenza di campionamento, ecc.), nonché qualsiasi informazione relativa ai valori che i dati stessi possono assumere (p. es., se le osservazioni al disotto di una certa soglia quantitativa non vengono registrate, ecc.).

Fornire tutte queste informazioni in una modalità che sia comprensibile ad un soggetto esterno alla PA che ha raccolto dei dati può essere faticoso e costoso. La soluzione? Per il futuro, pensate un po' di più all'eventualità che i dati possano essere riusati da terzi. Per il passato, chiarite che mancano dei metadati e che ci state lavorando, ma non fatevi paralizzare dall'assenza di una metadatazione perfetta: come al solito, procedere passo passo! (E considerate che, qualche volta, anche un approccio *quick & dirty* ha la sua utilità...).

No alla discriminazione tra umani e macchine!

Non si dovrebbe mai discriminare l'accesso ai dati tra umani e macchine. Questo principio può far sorridere e più di un lettore avrà pensato a qualche romanzo di fantascienza, ma la questione è rilevante ed è una questione di libertà (di impresa, ma anche di azione democratica). La necessità di **registrarsi su un portale**, per esempio, oppure la necessità di **accettare esplicitamente una licenza** e simili sono piccole noie per un essere umano, ma possono diventare ostacoli insormontabili per una macchina. E **questi ostacoli possono rendere i vostri dati difficili da trovare (perché i motori di ricerca non li vedono), oppure difficili da scaricare e aggiornare in automatico**, nell'ambito di applicazioni e servizi di terzi.

Se siete arrivati sin qui, potenzialmente ce l'avete fatta. Sapete il grosso di quel che serve a far partire un progetto open data! Ma ci sono ancora un certo numero di cose che potrebbero andar storte e che non abbiamo discusso, nonché alcune opportunità che abbiamo trascurato. La terza parte di questo libro bianco è proprio dedicata a queste sfide e ostacoli più insidiosi, nonché a far percepire – e possibilmente cogliere – alcune ulteriori opportunità per facilitare il riutilizzo e massimizzarne le ricadute positive.

3 Sfide, ostacoli e ulteriori opportunità

3.1 Le normativa sul riutilizzo dell'informazione del settore pubblico

È opportuno inaugurare questa sezione con un riferimento più puntuale e sistematico alla norme in materia di riutilizzo della PSI.

In Europa

La Direttiva 2003/98/CE del Parlamento europeo e del Consiglio, approvata il 17 novembre 2003 e pubblicata nella GUCE n. L 345 del 31 dicembre 2003, rappresenta a tutt'oggi il testo normativo di riferimento in tema di riutilizzo dell'informazione del settore pubblico nell'Unione Europea. Questa Direttiva si pone l'obiettivo di agevolare la *"creazione di prodotti e servizi a contenuto informativo, basati su documenti del settore pubblico, estesi all'intera Comunità, nel promuovere un effettivo uso, oltre i confini nazionali, dei documenti del settore pubblico da parte delle imprese private, al fine di ricavarne prodotti e servizi a contenuto informativo a valore aggiunto e nel limitare le distorsioni della concorrenza sul mercato comunitario"* (Considerando #25).

Naturalmente, la Direttiva non si applica indiscriminatamente a ogni dato detenuto dalle pubbliche amministrazioni, evitando dunque di pregiudicare diritti di terzi, tutela della sicurezza nazionale, segreto statistico o tutela della privacy.

La Direttiva PSI dispone, come principio generale, che gli Stati membri provvedano affinché, ove sia permesso il riutilizzo di documenti in possesso degli enti pubblici, questi documenti siano riutilizzabili a fini commerciali o non commerciali e resi disponibili, ove possibile, per via elettronica.

Tra i principali punti, la Direttiva prescrive che:

- gli Enti pubblici possano richiedere un corrispettivo in denaro per la messa a disposizione dei dati: in questo caso, le tariffe applicate non devono superare i costi di raccolta, produzione, riproduzione e diffusione dei documenti richiesti, maggiorati di un congruo utile sugli investimenti;
- le condizioni e le tariffe standard applicabili per il riutilizzo di documenti in possesso di enti pubblici vengano fissate in anticipo e pubblicate, ove possibile e opportuno, per via elettronica;
- gli Stati membri prevedano modalità pratiche per facilitare la ricerca di documenti disponibili per il riutilizzo, come elenchi di contenuti, di preferenza accessibili per via elettronica, dei documenti più importanti e dei portali collegati a elenchi di contenuti decentralizzati;
- siano vietati accordi di esclusiva per il riutilizzo dei dati, a meno che tali accordi non si rendano necessari in virtù di un evidente interesse pubblico.

Alla luce di una consultazione pubblica, svolta nel dicembre 2011, la Commissione Europea ha rilasciato una proposta di modifica della Direttiva PSI. Tale proposta di

modifica include:

- uno specifico riferimento alla messa a disposizione dei dati in formati *machine readable*;
- una regola base che individua nei soli costi marginali di riproduzione e disseminazione il tetto alla tariffazione dei dati, pur con possibili eccezioni;
- l'esplicita menzione, pur entro un regime specifico, dei dati relativi alle istituzioni culturali.

In Italia

L'attuazione italiana della direttiva comunitaria è avvenuta con il Decreto legislativo 24 gennaio 2006, n. 36, pubblicato nella G.U. del 14 febbraio 2006, n. 37. Il Decreto Legislativo 36/2006 è stato successivamente modificato dalla L. 96/2010.

Più recentemente, la cosiddetta Legge di semplificazione (Legge 4 aprile 2012, n. 35) riconosce il ruolo dei dati aperti, mentre la bozza di Agenda Digitale italiana promossa alla fine del 2011 e sottoposta a consultazione pubblica contiene un'ampia sezione dedicata agli open data.

A livello regionale

La Regione Piemonte è stata la prima ad aver predisposto specifici strumenti normativi riguardanti il riutilizzo dell'informazione pubblica. A seguito della Delibera di Giunta 31 - 11679 del 29 giugno 2009, che definiva le linee guida regionali per i processi di riutilizzo, associate alla definizione di licenze standard per tipologia di dati e categorie di utenza che definiscono le discipline d'uso dei dati regionali messi a disposizione, e della nuova versione di tale delibera (Delibera di Giunta regionale 36 - 1109 del 30 novembre 2010), dal 23/12/2011 è in vigore la Legge Regionale n. 24 sui dati aperti ("Disposizioni in materia di pubblicazione tramite la rete Internet e di riutilizzo dei documenti e dei dati pubblici dell'amministrazione regionale") mediante cui l'amministrazione regionale si vincola ad assicurare la disponibilità, la gestione, l'accesso, la trasmissione, la conservazione e la fruibilità dei dati in modalità digitale. Diventano quindi riusabili tutti i dati in possesso della Regione di tipo aggregato/anonimo o senza vincoli di privacy e che adottano come licenza standard la Licenza CC0 (o, in alternativa, CC BY).

Altre amministrazioni regionali si stanno dotando di analoghi dispositivi. La Legge regionale 14 luglio 2011, n. 9 della Regione Friuli Venezia Giulia ("Disciplina del sistema informativo integrato regionale del Friuli Venezia Giulia") fa della promozione della trasparenza secondo la metodologia degli open data uno dei suoi principali obiettivi. In altre Regioni sono in corso di approvazione, sotto forma di disegno o proposta di legge, prescrizioni normative sul tema degli open data: è il caso delle amministrazioni regionali di Lazio, Basilicata, Umbria, Lombardia, Sicilia, Puglia, Campania, e della Provincia di Trento.

Una mappa aperta delle leggi regionali in materia di open data, a cura di Ernesto Belisario, è disponibile al seguente indirizzo: http://g.co/maps/xhz59.

A livello cittadino

Dato il crescente interesse per le politiche open data, sarebbe difficile fornire un quadro esaustivo di tutte le iniziative a livello cittadino, quali le delibere adottate da giunte e consigli comunali. Può invece essere opportuno segnalare casi, quale quello piemontese, in cui vari comuni (per altro stimolati da un progetto sostenuto dall'ANCI) hanno aderito alle iniziative portate avanti dalla propria regione. In questo modo, le infrastrutture come i portali open data regionali vengono messe a disposizione di soggetti pubblici per i quali sarebbe stato inefficiente creare una propria autonoma piattaforma di pubblicazione dei dati.

3.2 Su costi, profitti e tariffe

In precedenza abbiamo illustrato come la migliore politica per massimizzare le opportunità del riutilizzo dei dati sia quella di limitare l'eventuale tariffa al costo marginale di messa a disposizione degli stessi (un costo che, in ambito digitale, è in sostanza pari a zero). Questo approccio permette la massima diffusione dei dati e, di conseguenza, degli effetti positivi generati. Nel rilasciare liberamente i propri dati, l'obiettivo della PA non deve essere infatti quello di "fare cassa", bensì di incentivare lo sviluppo di nuove applicazioni. Peraltro, l'attuale processo di revisione della Direttiva europea di riferimento va in questa direzione: salvo eccezioni motivate, il prezzo dei dati non può superare il costo sostenuto per metterne a disposizione un'unità aggiuntiva. È dunque possibile considerare, ad esempio la (minima) quota di ulteriore banda necessaria, o analoghi costi di natura tecnica, ma certamente non i costi fissi di realizzazione del servizio. Senza dimenticare che ogni pagamento implica non solo introiti, ma nuovi costi per la PA (legati alla gestione delle transazioni), costi che in alcuni casi possono rendere addirittura antieconomico il rilascio dei dati a titolo oneroso!

Un aspetto economico relativo a riutilizzo merita un commento specifico: la possibilità che il riutilizzo favorisca un determinato soggetto commerciale, apparentemente a spese dei contribuenti. [v. § 3.3.] La risposta, qui, è dunque del tipo "sì al riutilizzo commerciale, purché aperto a tutti". Ciò, val la pena esplicitarlo, è garantito dal fatto che le licenze standard di cui abbiamo parlato non sono mai esclusive; inoltre, le licenze esclusive sono esplicitamente scoraggiata (e normalmente vietate) dalla Direttiva PSI (e suo recepimento nazionale). Eventuali casi particolari, in cui malgrado l'assenza di licenze esclusive alcuni soggetti economici rafforzeranno la loro posizione dominante grazie alla PSI, saranno poi gestiti in via successiva, ad esempio tramite l'intervento dell'autorità antitrust.

GoogleTranslate ed il riutilizzo da parte di soggetti commerciali

Nel Maggio 2010, la parlamentare europea Evelyn Regner ha indirizzato alla Commissione Europea un'interrogazione riguardante l'utilizzo di documenti tradotti dell'UE da parte di Google, per lo sviluppo del suo software di traduzione.

Come noto, Google ha sviluppato il servizio di traduzione GoogleTranslate, liberamente disponibile online. Evelyn Regner chiedeva dunque: "**È la Commissione a conoscenza del fatto che l'azienda statunitense Google Inc. utilizza i documenti dell'UE tradotti in 23 lingue quale base per un servizio gratuito, perseguendo tuttavia un interesse economico?** Un siffatto servizio gratuito induce un numero sempre maggiore di utenti Internet in tutto il mondo a consultare la homepage di Google. Qual è la posizione della Commissione rispetto a tale riguardo?" Continuando poi: "**Come giustifica la Commissione ai contribuenti europei il fatto che per i servizi di traduzione siano stanziati milioni di euro e un'azienda privata statunitense ne tragga indirettamente beneficio?**"

Il Commissario Kroes ha risposto a nome della Commissione, confermando una piena consapevolezza dei fatti descritti. Nella risposta, il Commissario ha fatto addirittura riferimento ad altri usi, commerciali e non-commerciali, di altri documenti pubblicati in più lingue dalla Commissione, dal Parlamento e da varie istituzioni europee.

Il riutilizzo descritto è stato posto nel contesto della **Direttiva PSI (2003/98/EC), che ha precisamente lo scopo di incoraggiare questo tipo di attività e le loro ricadute (economiche e non), normalmente positive per la società**. Rispetto al caso specifico, poi, la Commissione ha ricordato che Google, pur essendo un operatore basato negli Stati Uniti, ha affiliate in più di una dozzina di paesi europei (inclusa l'Italia). Inoltre – e soprattutto – la Commissaria ha fatto osservare che **l'informazione in esame è disponibile a qualsiasi compagnia alle medesime condizioni e che i costi di traduzione, lungi dall'essere un aiuto a Google, sono spese inevitabili e connesse ai compiti delle istituzioni europee**. In conclusione, la Commissione non ha preso in considerazione l'imposizione di alcun limite aggiuntivo all'accesso e al riutilizzo dei documenti in questione.

Trattando di concorrenza, un altro caso merita una menzione. Immaginiamo che qualche riutilizzatore "consolidato" si lamenti di un'eventuale riduzione delle tariffe o in generale di una maggiore apertura (che faciliti l'ingresso di nuovi entranti, nonché suoi concorrenti), sostenendo magari che la sua possibilità di stare sul mercato potrebbe a quel punto essere minacciata. Se così accadesse, si avrebbe un netto indizio del fatto che quell'operatore basava la proprio attività su un posizionamento (quasi) monopolistico (o comunque da rendita di posizione), più che sulla creazione di servizi realmente innovativi. Altrimenti non avrebbe di che preoccuparsi. Di conseguenza, perché una PA dovrebbe farsi scrupoli ad aprire i dati, malgrado queste obiezioni?

I dati sulle imprese

Uno dei più vasti e articolati insiemi di dati gestiti dal settore pubblico è quello riguardante l'attività delle **imprese** operanti sul territorio italiano. Ogni azienda è infatti tenuta a condividere e aggiornare, sin dalla sua nascita, le principali informazioni in merito alla propria attività (ad esempio i propri dati di bilancio) presso la **Camera di Commercio** di competenza. Questi dati, insieme con i relativi documenti di origine, alimentano – su base nazionale – il Registro Imprese, gestito da Infocamere, società consortile informatica delle Camere di Commercio italiane. Questa banca dati presenta un particolare interesse, sia per i singoli operatori (tra i possibili esempi, un istituto di credito che voglia verificare lo stato patrimoniale di un cliente), sia per il sistema nel suo complesso: l'analisi dei dati, anche mediante l'incrocio con altre basi informative, costituisce una fotografia attendibile e aggiornata dello stato delle attività produttive in Italia.

Quello dell'informazione sulle imprese è ad oggi un **mercato** caratterizzato da dinamiche peculiari. Il rilascio dei dati da parte di Infocamere avviene secondo canali complementari tra loro: è possibile per i singoli cittadini interrogare puntualmente (a pagamento) il registro elettronico; l'accesso alla banca dati nel suo complesso è invece regolato secondo accordi pluriennali stipulati con operatori privati (riconosciuti da Infocamere come **distributori**, il cui numero è attualmente di 43) i quali forniscono consulenza al mercato; ulteriori accordi possono essere attivati mediante il servizio "Telemaco".

Nel caso di un settore consolidato come quello dell'informazione camerale (stimato valere almeno 500 milioni di euro), più che i modelli di licenza e

rilascio, è in sé la **configurazione di mercato** a determinare l'efficienza complessiva, ossia i benefici derivanti dalla condivisione dei dati. In particolare, il raggiungimento dell'**equilibrio** tra la necessità – per Infocamere – di sostenere la propria attività nel tempo e l'esigenza – nel mercato – di creare servizi intelligenti a vantaggio dei cittadini e delle imprese è strettamente collegato a una gestione accorta dei processi di apertura delle banche dati.

L'accesso libero e gratuito ai dati "grezzi" per chiunque, demandando invece al mercato la realizzazione di servizi complessi (tuttora richiesti e necessari), potrebbe rappresentare un'interessante opportunità da esplorare, a beneficio della collettività. Questo tipo di configurazione di mercato, tuttavia, richiederebbe di compensare il sistema camerale degli attuali introiti derivanti dalla "vendita" dei dati, stimabili nell'ordine di grandezza delle decine di milioni di euro (probabilmente tra 15 e 60, a seconda dell'effettivo impatto dell'apertura sugli introiti attuali).

3.3 Tutela dei dati personali: tra accesso e riutilizzo

Hic sunt leones. Questo è ciò che potete aspettarvi di trovar scritto, accanto alla voce "dati personali", nella *road-map* del tipico progetto open data. Senza dubbio, all'intersezione tra il terreno dei dati detenuti dal settore pubblico e quello dei dati personali, si trova un'area relativamente inesplorata e non priva di alcuni rischi. Questo è vero in particolare in Europa e in Italia, dove la tutela dei dati personali è particolarmente rigorosa. Ciò non significa che si tratti di un tipo di dati il cui riutilizzo sia escluso a priori; ma **può essere opportuno non affrontare il tema dati personali nella primissima fase di un progetto di apertura dei dati pubblici**, rimandando la complessità aggiuntiva introdotta dalla normativa in materia di protezione dei dati personali/privacy[11] ad una fase in cui si sia presa confidenza con l'approccio open data in generale.

Infatti, **la decisione di mettere a disposizione alcuni dati pubblici non deve (e non può, a norma di legge) mettere in discussione il diritto dei cittadini alla tutela dei loro dati personali**: la normativa sul riutilizzo dei dati fa infatti completamente salva la normativa sui **dati personali** con particolare riferimento alle regole di sicurezza e misure minime previste dall'allegato B al Codice della Privacy a garanzia dei dati dei cittadini e delle imprese. Dunque, quando un data set contiene dei dati personali, è bene consultare un legale esperto di privacy per valutare, verificare e stabilire con apposito provvedimento interno quali dati presenti nelle banche dati siano conoscibili da chiunque e a quali condizioni, e quando si è in presenza di dati che rientrano tra i cosiddetti dati sensibili, l'ipotesi più plausibile è che l'accesso e il riutilizzo siano semplicemente possibilità da scartare. È quindi opportuno avere un'idea di quali dati siano considerati personali comuni e di quali, tra questi, siano considerati dati sensibili[12].

I dati personali sono una classe molto ampia, comprendente **qualsiasi informazione concernente una persona fisica identificata o identificabile**[13]. Alcuni **esempi** tipici

11 I cultori della materia ci perdoneranno l'uso del termine "privacy" come abbreviazione per "tutela dei dati personali", ricordando che lo stesso URL del Garante per la Protezione dei Dati Personali è http://www.garanteprivacy.it/.

12 Le definizioni sono contenuto nell'art. 4 del Codice in materia di protezione dei dati personali (D.Lgs n. 196 del 2003): http://www.garanteprivacy.it/garante/doc.jsp?ID=1311248.

13 Questa definizione è tratta dalla Direttiva 95/46/CE riguardante il

comprendono: il nome e cognome, l'indirizzo di residenza, il codice fiscale, l'indirizzo mail, il numero di targa o il numero di telefono, ma anche l'immagine fotografica di una persona e la registrazione della sua voce. Inoltre, quando si parla di persona identificabile, lo si fa in senso abbastanza ampio e anche indiretto, ad esempio "mediante riferimento ad un numero di identificazione o ad uno o più elementi specifici caratteristici della sua identità fisica, fisiologica, psichica, economica, culturale o sociale".

Vi sarete certamente resi conto di quanti dati, il cui riutilizzo sarebbe potenzialmente utile, rientrino nella categoria "dati personali". Fortunatamente, a certe condizioni, questi dati possono essere resi accessibili e il loro riutilizzo può essere lecito. Prima di fornire una rapida introduzione a queste condizioni, tuttavia, è bene chiarire che ci sono una serie di dati che è opportuno – almeno in prima approssimazione – considerare assolutamente off-limits ed esclusi da qualsiasi ipotesi di accesso e riutilizzo. I **dati sensibili** sono quei dati personali che si riferiscono alla sfera più intima del soggetto, **"idonei a rivelare l'origine razziale ed etnica, le convinzioni religiose, filosofiche o di altro genere, le opinioni politiche, l'adesione a partiti, sindacati, associazioni od organizzazioni a carattere religioso, filosofico, politico o sindacale, nonché i dati personali idonei a rivelare lo stato di salute e la vita sessuale"**[14]. Alcuni **esempi** comprendono l'elenco dei titoli relativi alle condizioni di salute nei concorsi pubblici (p. es. soggetti diversamente abili); l'iscrizione ad un partito politico e ad un'associazione sindacale, l'elenco degli studenti che seguono o non seguono le ore di religione a scuola o i file di log di navigazione Internet su siti erotici, politici, sindacali. Ma occorre anche prestare attenzione a casi in cui un dato che può sembrare semplicemente personale – ad esempio l'indirizzo mail – diventa sensibile, poiché rivela l'adesione ad una determinata associazione, sindacato o partito, p. es. alberto.rossi@cgil.it o mario.verdi@pd.it.

Esiste una categoria di dati detti "semisensibili", il cui trattamento "presenta rischi specifici". In merito, il Codice della privacy (art. 17) stabilisce che "[i]l trattamento dei dati diversi da quelli sensibili e giudiziari che presenta rischi specifici per i diritti e le libertà fondamentali, nonché per la dignità dell'interessato, in relazione alla natura dei dati o alle modalità del trattamento o agli effetti che può determinare, è ammesso nel rispetto di misure e accorgimenti a garanzia dell'interessato, ove prescritti. [Tali misure e accorgimenti] sono prescritti dal Garante [...], anche a seguito di un interpello del titolare."

Un esempio di "trattamento che presenta rischi specifici" è la pubblicazione online dell'elenco di genitori che non pagano le rette della scuola o dell'asilo. Un caso di questo genere permette anche di comprendere perché vi siano alcune modalità di trattamento ammissibili e altre più delicate: ad esempio, per una persona dal nome e cognome poco diffusi e senza un'autonoma presenza online, l'elenco dei genitori in arretrato sulle rette potrebbe essere l'unica

trattamento dei dati personali e la libera circolazione degli stessi. Si noti che, secondo la normativa italiana, la definizione è ancora più ampia e ricomprende "qualunque informazione relativa a persona fisica, persona giuridica, ente od associazione, identificati o identificabili" (art. 4 Codice in materia di protezione dei dati personali). La pubblicazione di dati relativi a persone giuridiche dovrà essere particolarmente cauta, nel caso dal dato in questione si possano desumere informazioni relative a persone fisiche.

14 Il Codice della Privacy ha previsto rigorose garanzie per il trattamento di dati sensibili: in ambito privato i dati sopra citati possono essere oggetto di trattamento solo con il consenso scritto dell'interessato e previa autorizzazione del Garante per la protezione dei dati personali; mentre il trattamento di dati sensibili da parte degli enti pubblici è ammesso solo se autorizzato da espressa disposizione di legge nella quale sono specificati i tipi di dati che possono essere trattati, le operazioni eseguibili e le finalità di rilevante interesse pubblico perseguite (art. 20) Gli enti pubblici devono trattare i dati sensibili in modo da prevenire violazione dei diritti, delle libertà fondamentali e della dignità dell'interessato e devono trattare solo i dati indispensabili per svolgere le attività istituzionali che non possono essere adempiute mediante il trattamento di dati anonimi o di dati di diversa natura.

presenza online. In questo caso, dal punto di vista di chi cercasse il suo nome con un motore di ricerca (ad esempio, un reclutatore prima di un colloquio di lavoro), la persona in esame sarebbe "connotata" solo come "morosa"... una situazione che presenta un certo rischio per la dignità dell'interessato).

Gli enti pubblici sono esonerati dal richiedere ai soggetti il consenso al trattamento dei dati personali ma possono trattare i dati personali soltanto per lo svolgimento delle funzioni istituzionali (v. art. 18 del Codice della privacy). L'ente pubblico, anche in un'ottica di apertura dei dati pubblici, deve individuare ed esplicitare con esattezza con un atto interno le funzioni istituzionali svolte e di cui sia effettivamente titolare in base all'ordinamento di riferimento (leggi, statuti e regolamenti). In presenza di una richiesta di riutilizzo da parte di imprese e privati, l'ente deve valutare se le finalità di riutilizzo prospettate possano essere considerate compatibili con le finalità istituzionali originarie e delineate nel documento interno.

Gli ostacoli pratici all'accessibilità hanno rappresentato, tradizionalmente, un elemento del compromesso tra diritto all'accesso[15] e diritto alla riservatezza. Se, per ottenere un certo dato, è necessario recarsi in un dato ufficio, magari nel limitato orario di apertura al pubblico, e magari pagare un corrispettivo per avere una fotocopia di un documento contenente il dato, allora quel dato, benché potenzialmente accessibile a chiunque in modo trasparente, è protetto da una sorta di "opacità". Oggi, la disponibilità in formato digitale dei dati pubblici diminuisce, sin quasi ad annullarla quella che viene a volte definita "privacy by obscurity" ("riservatezza tramite l'oscurità"). L'esempio che segue mostra la previsione di ostacoli tecnici alla libera fruizione di dati pubblici, al fine di ricostruire per via tecnologica un certo grado di "privacy by obscurity".

15 La disciplina sul riutilizzo dei dati pubblici fa salve le disposizioni in materia di accesso ai documenti amministrativi: non si applica la direttiva nei casi in cui cittadini o le imprese debbano dimostrare, in virtù del regime di accesso, di avere un particolare interesse all'ottenimento dell'accesso.

CV dei dirigenti pubblici e Robots.txt

Il provvedimento del 16 luglio 2009 del Garante della Privacy fa riferimento alla pubblicazione online dei CV di dirigenti pubblici. Il Garante raccomanda ("ritiene preferibile") alle PA l'uso di strumenti che evitino o limitino l'indicizzazione dei dati da parte dei normali motori di ricerca generalisti, cosicché siano assicurati "accessi maggiormente selettivi e mirati ai soli dati personali coerenti con le finalità sottese alla pubblicazione, e cioè con l'esigenza di assicurare, presso ciascuna Amministrazione, la trasparenza delle informazioni di cui all'articolo 21, comma 1, della legge n. 69 del 2009". Il Garante indica inoltre i principî giuridici in materia di protezione dei dati personali che determinano questa scelta: "per rispettare il principio di pertinenza e non eccedenza dei dati personali, il trattamento dei dati oggetto di pubblicazione non può riguardare informazioni che non siano funzionali a tale finalità di trasparenza".

L'informazione giuridica (e le sentenze, che contengono dati personali)

I testi di legge vigenti, le sentenze emanate dai Tribunali, nonché le prassi amministrative prodotte

dalle Autorità, compongono l'informazione giuridica del paese. Per avere un'idea dell'ammontare di informazioni di cui parliamo, basti pensare che in un anno sono oltre 50mila le sentenze prodotte dalla sola Cassazione. Oltre ai professionisti dell'ambito giuridico, e ai cittadini in generali, sono molti i soggetti che potrebbero aver interesse a conoscere alcuni di questi dati, ad esempio le compagnie assicurative per le decisioni relative ai sinistri.

Di recente sono stati compiuti alcuni importanti passi avanti rispetto alla apertura delle banche dati giuridiche. L'intero corpo delle leggi vigenti in Italia è consultabile (da qualche anno gratuitamente) presso il portale "Normattiva" (www.normattiva.it) gestito dall'Istituto Poligrafico e Zecca dello Stato. Risulta quindi relativamente agevole, per chiunque, verificare le attuali disposizioni di legge rispetto a una particolare disciplina di interesse.

Le sentenze della Cassazione sono invece consultabili – corrispondendo un diritto di segreteria – in cartaceo presso le sedi di competenza; alcune di queste, ritenute di particolare interesse, vengono riassunte e divulgate (a pagamento) in formato elettronico mediante il Centro Elettronico di Documentazione della stessa Cassazione. In linea generale, tuttavia, la comprensione (per i non addetti ai lavori) dei testi di giurisprudenza necessita di particolari strumenti interpretativi. A valle di un lavoro di catalogazione, commento e collegamento con altre banche dati giuridiche, l'insieme delle sentenze rientra infatti nell'offerta, rivolta principalmente a liberi professionisti, di alcuni operatori specializzati nell'editoria giuridica. Si tratta di editori, come Zanichelli, Giuffrè, Wolters Kluwer Italia, dotati di una forte componente reputazionale costruita negli anni a partire dalla realizzazione di prodotti editoriali, disponibili ormai anche in formato elettronico.

Anche nel rilascio delle sentenze di massimo grado (Cassazione) sembra dunque possibile esplorare nuovi modelli di apertura. In primo luogo per ciò che riguarda il formato delle stesse (il trasferimento in digitale delle sentenze rappresenta un costo per gli editori che potrebbe essere re-investito nell'ulteriore miglioramento dei propri prodotti). Ma anche nella messa a disposizione, in digitale e senza oneri, delle sentenze relative a specifiche discipline, caratterizzate da un numero ridotto e "marginali" nell'ambito della complessa offerta editoriale attuale, ma di interesse per alcuni utenti finali, siano essi liberi professionisti o ricercatori.

Il quadro normativo di riferimento

Le sentenze sono atti pubblici, pertanto non ci sono limiti normativi generali alla loro apertura.

Come ha ricordato il Garante per la protezione dei dati personali nelle *Linee guida in materia di trattamento di dati personali nella riproduzione di provvedimenti giurisdizionali per finalità di informazione giuridica* (2010), "la diffusione dei provvedimenti giurisdizionali costituisce fonte preziosa per lo studio e l'accrescimento della cultura giuridica e strumento indispensabile di controllo da parte dei cittadini dell'esercizio del potere giurisdizionale. Il Codice in materia di protezione dei dati personali favorisce la più ampia diffusione delle sentenze e degli altri provvedimenti dell'Autorità giudiziaria per i quali sia stato assolto, mediante il deposito nella cancellerie e nelle segreterie giudiziarie, l'onere della pubblicazione previsto dalle disposizioni dei codici di procedura civile e penale". I riferimenti normativi sono costituiti nello specifico dagli artt. 51 e 52

d. lgs. 196/2003, nonché dall'art. 56, cc. 2 e 2-bis, d. lgs. 82/2005 (Codice dell'amministrazione digitale). Fatte salvi alcuni limiti specifici e residuali di cui all'art. 52, cc. 2 e 5, d. lgs. 196/2003, "è ammessa la diffusione in ogni forma del contenuto anche integrale di sentenze e di altri provvedimenti giurisdizionali" (art. 52, c. 7 Codice in materia di protezione dei dati personali).

In caso neppure l'interazione con un legale riesca a chiarire i dubbi rispetto all'accesso e riutilizzo di un determinato insieme di dati personali e nell'ambito di una certa finalità, ricordate infine che è possibile interpellare direttamente il Garante Privacy.

Il caso di Voglioilruolo.it

Voglioilruolo nasce da un bisogno reale: l'enorme (e disomogenea) quantità di informazioni disponibili in Rete in merito alle "graduatorie ad esaurimento", utilizzate dal MIUR per assegnare incarichi di insegnamento a tempo indeterminato o determinato. A partire dalla raccolta e dall'elaborazione di dati relativi a oltre 20mila graduatorie, *Voglioilruolo* è un servizio che si propone di assicurare una maggiore fruibilità di tali dati, garantendo una visualizzazione compatta di tutte le informazioni relative a ogni docente (come ad esempio le province nelle quali è iscritto, o i concorsi a cui è abilitato a partecipare). Da cui la possibilità di calcolare, date le proprie caratteristiche, la provincia che assicura la maggiore probabilità di chiamata, o ancora di navigare le graduatorie con funzioni di ricerca avanzate, entro un modello di tipo *freemium*. Naturalmente, le informazioni trattate da *Voglioilruolo* contengono dati personali. A partire dalla ritenuta impossibilità di poter rilasciare a ciascun iscritto nelle graduatorie (circa 300.000 unità) l'informativa individuale per il trattamento dei dati personali, i responsabili di *Voglioilruolo* hanno richiesto al Garante della Privacy l'esonero da tale adempimento. Tale richiesta è stata accolta in virtù dell'utilità del progetto, riconosciuta dallo stesso Garante.

3.2.1 L'informativa

Una corretta informativa relativa alla privacy è un elemento essenziale del trattamento dei dati personali. Inoltre, la presenza di un'informativa chiara – e che prenda esplicitamente in considerazione il fatto che i dati saranno pubblicati su Internet e potenzialmente riutilizzati – può essere uno strumento prezioso per gli stessi riutilizzatori.

In particolare, può essere utile abbinare ai dataset contenenti dati personali un'indicazione delle ragioni (e delle norme) in virtù delle quali questi dati sono stati raccolti e pubblicati, giacché, in mancanza di una tale indicazione, ogni terza parte che volesse riutilizzare dati personali al di fuori di questo confine rischierebbe di dover ottenere il preventivo consenso di tutti gli interessati (e questo chiaramente a prescindere dalle licenze di proprietà intellettuale associate ai data base in questione).

Gli enti pubblici, in qualità di titolari del trattamento dei dati, prima di iniziare la raccolta dei dati stessi devono rilasciare agli interessati (i soggetti a cui si riferiscono i dati) uno specifico documento informativo (c.d. "informativa")

al fine di illustrare le puntuali caratteristiche minime del trattamento indicati in modo tassativo dall'art. 13 del Codice della Privacy: indicazione delle finalità e delle modalità del trattamento a cui sono destinati i dati; la natura obbligatoria o facoltativa del conferimento dei dati; le conseguenze di un eventuale rifiuto di rispondere; i soggetti o le categorie di soggetti ai quali i dati personali possono essere comunicati o che possono venirne a conoscenza; i diritti sui propri dati indicati dall'art. 7 del Codice della Privacy (diritto di accesso sui propri dati, di intervento e di opposizione). Risulta pertanto strategico per l'ente riesaminare le informative privacy rilasciate e provvedere all'eventuale integrazione comprensiva della possibilità del riutilizzo dei dati: è infatti necessario che gli interessati siano informati della possibilità di opporsi a trattamenti sia pure legittimi[16].

3.2 2 L'anonimizzazione

Come abbiamo detto, i dati personali sono delicati da gestire. Fortunatamente, è sempre possibile adottare una soluzione tecnica, che rende possibile la pubblicazione di un dato, semplicemente eliminandone la natura personale. Infatti, **un dato è personale finché resta riconducibile ad una persona fisica e questa riconducibilità può venir meno grazie a varie tecniche di anonimizzazione**[17]. Tra queste tecniche rientra anche l'**aggregazione** di più dati personali, ad esempio calcolando una media per quartiere, città, fascia d'età, ecc.

A scanso di equivoci, l'anonimizzazione è una **soluzione efficace, ma radicale, poiché comporta la perdita di molte informazioni** potenzialmente utili. Ad esempio, se si anonimizza aggregando dei dati, il criterio di aggregazione rifletterà uno o più punti di vista specifici. Pensate all'informazione relativa alle convinzioni religiose di un individuo. Per certi fini, ad esempio per decidere se aprire una macelleria kosher in una certa area, sarà utile avere l'informazione per quartiere; ma per altri fini, ad esempio per aprire un asilo privato, sarà utile avere l'informazione anche (o forse soprattutto) per età (anche a costo di conoscerla solo per circoscrizione e non per quartiere). Mentre ad altri fini interesserà lo stesso dato, ma diviso per fasce di reddito. Ma è inevitabile che l'aggregazione e l'anonimizzazione facciano allontanare dal dato grezzo, poiché rimuovendo l'informazione sull'identità dell'individuo si elimina il punto di contatto, la chiave comune tra i dati menzioni prima: la riconducibilità al singolo individuo che ha una certa convinzione religiosa, un certo reddito, una certa età e vive in un certo luogo.

Il segreto statistico

Sono esclusi dall'applicazione della normativa sul riutilizzo, i documenti connessi alla tutela del segreto statistico, quali disciplinati dal decreto legislativo 6 settembre 1989, n. 322.

Il sopracitato D.Lgs. 6 settembre 1989, n. 322, chiarisce che "[i] dati elaborati nell'ambito delle rilevazioni statistiche comprese nel programma statistico nazionale sono patrimonio della collettività e vengono distribuiti per fini di studio e di ricerca a coloro che li richiedono [...]". La legge, dunque, riconosce uno

16 V. art. 14 della direttiva 95/46; v. art. 7 del Codice della privacy che prevede: L'interessato ha diritto di opporsi, in tutto o in parte: a) per motivi legittimi al trattamento dei dati personali che lo riguardano, ancorché pertinenti allo scopo della raccolta; b) al trattamento di dati personali che lo riguardano a fini di invio di materiale pubblicitario o di vendita diretta o per il compimento di ricerche di mercato o di comunicazione commerciale.

17 Ai sensi dell'art. 4 del Codice della privacy, il dato anonimo", è il "il dato che in origine, o a seguito di trattamento, non può essere associato ad un interessato identificato o identificabile".

dei presupposti fondamentali dell'open data (il fatto che i dati siano patrimonio della collettività), anche se pare limitare lo scopo del riutilizzo ad alcuni fini specifici (studio e ricerca).

Inoltre, l'art. 9 della stessa norma stabilisce anche alcune "Disposizioni per la tutela del segreto statistico". L'opportunità di simili norme appare chiara, qualora si consideri l'esistenza di uno specifico obbligo a fornire alcune informazioni (potenzialmente sensibili o tali da permettere la ricostruzione di informazioni sensibili), ad esempio in occasione di un censimento. Al fine di ottenere risposte veritiere da parte di cittadini e imprese, dunque, è ragionevole imporre limitazioni al riutilizzo delle informazioni stesse.

In particolare, "[i] dati raccolti nell'ambito di rilevazioni statistiche comprese nel programma statistico nazionale [...] non possono essere esternati se non in forma aggregata, in modo che non se ne possa trarre alcun riferimento relativamente a persone identificabili e possono essere utilizzati solo per scopi statistici."

Al fine di prevenire eventuali operazioni illecite di *data mining*[18], la legge stabilisce anche che "[i]n ogni caso, i dati non possono essere utilizzati al fine di identificare nuovamente gli interessati." D'altro canto, il segreto statistico non può essere arbitrariamente esteso e utilizzato come scusa: come stabilito dallo stesso articolo, infatti, "non rientrano tra i dati tutelati dal segreto statistico gli estremi identificativi di persone o di beni, o gli atti certificativi di rapporti, provenienti da pubblici registri, elenchi, atti o documenti conoscibili da chiunque."

18 Per data mining, si intendono le attività di estrazione eseguita in modo automatizzato di informazione implicita e nascosta, a partire da grandi quantità di dati già strutturati, per renderla disponibile e direttamente utilizzabile.

Naturalmente, infine, l'aggregazione ha come obiettivo la tutela della riservatezza, non l'opacità o l'oscurità delle procedure seguite o del tipo di dati originariamente raccolti. Per questo, e sia pure solo in base ad una richiesta motivata, il medesimo decreto prevede che possano essere distribuite anche collezioni campionarie di dati elementari, opportunamente resi anonimi.

3.3 Controlli, abusi e responsabilità

3.3.1 Obblighi o raccomandazioni? Registrazione obbligatoria degli utenti? "No, grazie."

È comune (e corretto) che le amministrazioni pubbliche vogliano conoscere chi usa i loro dati, perché e con quali risultati. Ciò aiuta le amministrazioni stesse a supportare il riutilizzo e migliora l'*accountability* dei progetti open data. Tuttavia, questo non significa che sia una buona idea obbligare gli utenti dei portali open data a registrarsi e/o a notificare l'uso che fanno dei dati. Non di rado, infatti, un'amministrazione può ottenere una buona gamma di informazioni combinando le statistiche del sito (accessi, download, etc.), i risultati di eventuali sondaggi facoltativi o simili ed altri meccanismi che creino un incentivo a condividere informazioni (banalmente, una sezione per segnalare applicazioni e riusi).

Ovviamente, anche la registrazione degli utenti è un'opzione, finché sia di tipo facoltativo e non condizioni l'accesso e le possibilità di riutilizzo dei dati.

3.2 2 Responsabilità (dai disclaimer più prudenti alla certificazione del dato)

Nell'ambito delle licenze standard relative ad eventuali diritti d'autore o connessi contenuti nei dati messi a disposizione del pubblico (di cui si è già raccomandato l'uso nei progetti sui dati aperti) è prassi consolidata e generalmente consigliabile inserire clausole di limitazione e/o esenzione da responsabilità. Per esempio, si chiarisce normalmente che l'oggetto della licenza (il database o le opere) è fornito "nello stato in cui si trova" o "così com'è" (in inglese "as is") e si specifica di non fornire "alcuna garanzia di alcun tipo, espressa o implicita", spesso escludendo espressamente garanzie rispetto alla "idoneità per un particolare scopo" o alla "assenza di vizi occulti o di altri difetti", "esattezza, presenza od assenza di errori", il tutto "nella più ampia misura permessa dalla legge applicabile". In sostanza l'impiego dei dati è "a rischio e pericolo" del riutilizzatore: in queste ipotesi il referente normativo è l'art. 1229 c.c..

Insomma, chi voglia tutelarsi dall'assunzione di eccessive responsabilità ha molti strumenti standard a propria disposizione (o può ispirarsi ad essi, nel caso voglia – a proprio rischio e pericolo, e potenzialmente a discapito delle possibilità di effettivo riutilizzo – avventurarsi nella stesura di una licenza "su misura"). La motivazione di questa scelta potrebbe essere quella di evitare che la P.A. che mette a disposizione per il riutilizzo le informazioni possa essere convenuta in giudizio a titolo di responsabilità contrattuale (se vogliamo "leggere" queste licenze come contratti) od a titolo di responsabilità extracontrattuale in base agli artt. 2043 ss. c.c. (nel caso opposto).

Esclusione di ogni garanzia e responsabilità? Forse non è questo che vuole la PA!

Non sempre le pubbliche amministrazioni vogliono escludere ogni responsabilità. Ad esempio, Regione Piemonte utilizza sul proprio portale open data (dati.piemonte.it) delle licenze libere Creative Commons (in particolare la licenza/liberatoria CCo, la più ampia tra le licenze aperte), che normalmente escluderebbero qualsiasi garanzia per l'utilizzatore rispetto al fatto che il licenziante sia titolare esclusivo di diritti sulle opere e sui prodotti culturali e sia quindi legittimato a disporne (il dichiarante quindi "declina ogni responsabilità per il diritto al compenso che altre persone possono rivendicare [... e...] declina inoltre ogni responsabilità derivante dall'ottenimento di tutti i necessari consensi, permessi o altri diritti richiesti per l'utilizzo [...])". L'idea alla base di queste licenze, infatti, è "io, titolare, ti permetto di utilizzare liberamente qualcosa su cui ho dei diritti: per quel che riguarda i miei diritti, dunque, fai pure ciò che vuoi, nei limiti della licenza, ma non puoi chiedermi di garantirti che qualcun altro non accampi pretese sui medesimi dati". Data la propria natura di ente pubblico, tuttavia, Regione Piemonte ha deciso di non volersi "lavare le mani" rispetto ad un punto così importante, decidendo di premettere una propria nota alla licenza adottata: "La titolarità piena ed esclusiva di codesta banca dati, è della Regione Piemonte (Licenziante), ai sensi della L. 633/41 e s.m.i.". Questa scelta, aumentando il grado di fiducia dei riutilizzatori rispetto al fatto di avere un dato completamente ed assolutamente libero, non può che essere accolta con favore e rappresenta sicuramente una buona pratica da seguire, laddove sia possibile.

Una scelta ancor più favorevole al riutilizzatore, infine, può riguardare la "certificazione" vera e propria della qualità dei dati messi a disposizione. A tal riguardo, è opportuno osservare che lo standard implicito nel campo dei dati aperti è "non ti offro alcuna garanzia sulla qualità del dato, ma sappi che lo considero abbastanza buono da usarlo internamente per i miei fini istituzionali". In questo contesto, potrebbe essere interessante garantire in modo esplicito (anziché implicito) che il dato messo a disposizione per il riutilizzo è il medesimo che la PA usa internamente[19].

3.4 Si può fare di più (tecnicamente)... dati aperti a cinque stelle!

"Dati grezzi, ora!" è ormai la nostra parola d'ordine. Speriamo di avervi convinto che si tratti di un passo importante e, se preso sul serio, anche di un passo sufficiente a creare grandi occasioni per l'innovazione. Esplicitando i complementari aspetti giuridici, abbiamo anche usato lo slogan "Dati grezzi e aperti, ora!".

Premesso che l'importante resta cominciare, dal punto di vista tecnico si può fare di più che pubblicare i dati "così come sono". Il ricercatore considerato il "padre" del World Wide Web (e anche uno dei padri del movimento open data), Tim Berners-Lee, propone di classificare i progetti open data secondo il seguente modello "a stelle".

19 Qualora, poi, un utente desiderasse un livello di certificazione di qualità superiore a quello che la PA considera sufficiente per se stessa, si entrerebbe sicuramente nell'ambito dei servizi a valore aggiunto a pagamento: uno scenario di questo genere esula dall'ambito di questo libro bianco, ma possiamo almeno menzionare che la scelta di una tariffa ottima per questo tipo di servizio, nonché la concorrenza con eventuali "certificatori" privati, potrebbero porre problemi di una certa complessità.

Il Modello delle stelle di Berners-Lee[20]:

★ Mettete a disposizione i vostri dati sul Web (in qualsiasi formato, "così come sono").

★★ Metteteli a disposizione come dati strutturati (p. es. in formato Excel, anziché come immagini scannerizzate di una tabella).

★★★ Utilizzate formati non-proprietari/aperti (p. es. CSV anziché Excel).

★★★★ Usate degli URL (stabili) per identificare gli oggetti, così che le persone possano creare puntatori alla vostra roba (p. es. http://en.wikipedia.org/wiki/Uniform_Resource_Locator).

★★★★★ Collegate i vostri dati ai dati di altre persone, per fornire un contesto (vedere la scheda Linked Data).

Secondo Berners-Lee, ma anche secondo la maggior parte dei commentatori ed esperti del settore, gli open data ben fatti, dunque, sono collegati tra loro, ovvero sono anche "linked data". I dati a 5 stelle, i linked open data, sono certamente più impegnativi da mettere a disposizione rispetto ai dati "così come sono", ma rispettare questi standard elevanti potrebbe far fare un enorme salto di qualità ai riutilizzatori.

20 Vedere: http://www.w3.org/DesignIssues/LinkedData.html. Ci sono versioni simili, ma non del tutto identiche, p. es. http://roessler.posterous.com/government-data-done-well-the-eu-digital-agen.

Linked Data[21]

Si è già accennato in precedenza che la mescolabilità dei dati è uno dei vantaggi più importanti del modello open data. I dati, se isolati, hanno poco valore; viceversa, la loro utilità aumenta sensibilmente quando diversi insiemi di dati, prodotti e pubblicati in modo indipendente da diversi soggetti, possono essere incrociati liberamente da terze parti.

Il riutilizzo è facilitato al massimo quando è banale collegare i dati tra loro, stabilendo un link diretto se e quando i dati (eventualmente provenienti da diverse sorgenti) si riferiscono a oggetti identici o comunque relazionati tra loro. Tale collegamento diretto si manifesta come la possibilità di "saltare" da un dataset all'altro, ad esempio quando si vuole accedere a dati (come i dettagli su una particolare entità, p. es. la città di Ancona – vedi sotto) che non si posseggono all'interno.

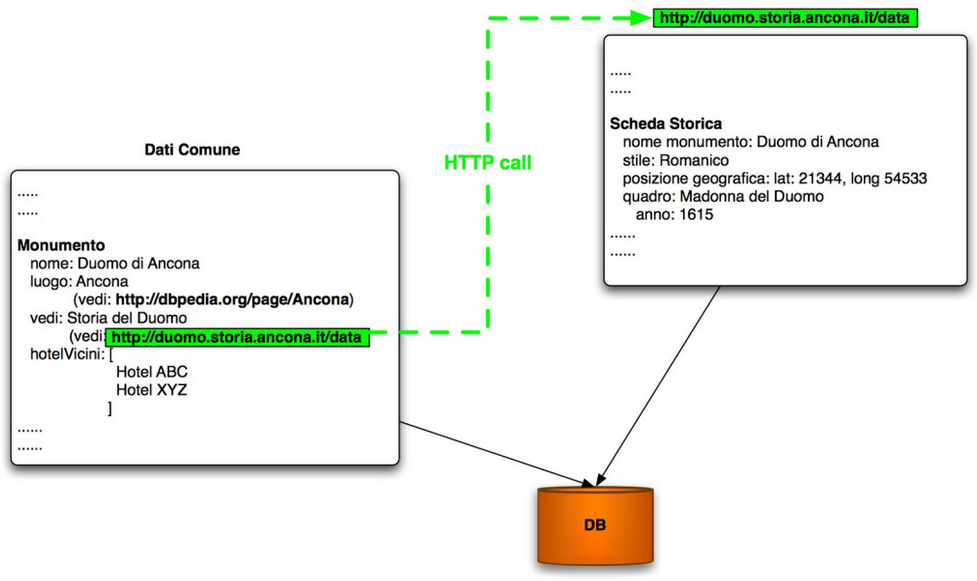
Supponiamo per esempio di avere, da una parte, amministrazioni locali che pubblicano dati aperti relativi ai monumenti storici e agli hotel che si trovano nelle vicinanze di quei monumenti; dall'altra, Sovrintendenze ai beni culturali che pubblicano dati dettagliati sui monumenti, gli artisti e i periodi storici, e sui quadri esposti nei musei o nei palazzi. Combinare i due dataset potrebbe essere di grande utilità, ad esempio per offrire un servizio personalizzato sugli itinerari in base agli interessi culturali specifici di un turista. Per fare questo, se i dati non sono "collegati" (linked) occorre in qualche modo creare questi link, processando i dati a mano o attraverso algoritmi ad hoc. Questo processo può non essere banale e sicuramente è una barriera al riutilizzo organico dei dati.

Nei cosiddetti "linked data", questi collegamenti e relazioni tra le entità descritte nei dataset sono già espliciti. Il Resource Description Framework (RDF) è lo standard di riferimento per la codifica, lo scambio e il riutilizzo di dati collegati in modo da essere massimamente interoperabili. In RDF i nomi sono Uniform Resource Identifier (URI) ed un tipo di URI particolare, molto utilizzato, è lo Uniform Resource Locator (URL). Gli URL sono familiari a chiunque usi il Web: sono semplicemente gli indirizzi che compaiono nella barra di navigazione del browser e cominciano con **"http://"**. Nel caso dei linked data, tale rappresentazione è una descrizione della risorsa stessa in forma di RDF.

Per esempio se parlo di Barolo, inteso come Comune, occorre stabilire un collegamento con la pagina http://it.dbpedia.org/resource/Barolo_(Italia) piuttosto che con quella del vino Barolo http://it.dbpedia.org/resource/Barolo.

Figura 1: il meccanismo di risoluzione degli URL permette di collegare un dataset all'altro consentendo, ad esempio, di mescolare i dati in un terzo database.

21 Questa scheda riutilizza liberamente materiale fornito da Christian Morbidoni e Francesca Di Donato di LinkedOpenData.it.



Linked Data Cloud

Il precedente esempio è solo uno degli scenari possibili in cui i linked data possono favorire l'interoperabilità tra dataset. Le possibilità sono infinite se pensiamo alla vasta quantità di linked open data già presenti sul Web. DBPedia.org, per esempio, espone una grande porzione di dati di Wikipedia.org come linked data, mentre Geonames.org offre descrizioni RDF di entità geografiche. Linkeddata.org fornisce un quadro dello stato corrente della "Linked data cloud", e mostra un ecosistema di database interconnessi in rapida crescita. Da pochissimo esiste anche it.dbpedia.org, che raccoglie e struttura dati provenienti dalle pagine Wikipedia in italiano.

Ciascuno di questi database fornisce dati che possono essere linkati dal vostro dataset, e che possono aggiungere valore ai vostri dati.

Naturalmente, i linked open data "ben fatti" sono quelli che usano formati standardizzati, interoperabili e machine-readable, al fine di rendere possibile la massima integrazione tra fonti diverse di dati (o diversi "silos informativi", se preferite). Per fare ciò servono standard, non necessariamente imposti centralmente, ma comunque condivisi o realizzati in modo tale da miminizzare gli sforzi per raggiungere l'interoperabilità.

Alcuni di questi standard sono rappresentati da dizionari come Dublin Core e Friend Of A Friend .

Tuttavia, ancora una volta, val la pena mettere in guardia contro la chimera rappresentata dal "progetto perfetto". Quando è possibile collegare i propri dati a semplici dizionari standard di riferimento come quelli menzionati sopra, è sicuramente opportuno farlo. Lo stesso vale per l'uso di metadati che rappresentino lo standard in un certo settore scientifico o disciplinare (p. es. lo standard SDMX nel campo dei dati statistici). Ma sarebbe alquanto pericoloso e controproducente ritardare la messa a disposizione di alcuni dati, attendendo di aver individuato gli standard di riferimento ideali per "linkarli" in modo opportuno. Allo stesso modo, l'esperienza di molti informatici che hanno collaborato con le pubbliche amministrazioni in passato pare concorde nel mettere in guardia rispetto a progetti che puntino a realizzare "ontologie" della PA capaci di rappresentare in modo standard la pletora di complesse attività, ruoli ed entità che le PA trattano o con cui entrano in contatto.

Insomma, ottimo puntare alle 5 stelle del linked data fatto ad arte, ma a volte può essere accettabile far riferimento a dizionari e metadati con validità solo "locale": se emergerà uno standard, qualcuno (comprese la stessa PA che ha messo a disposizione i dati in un primo momento) potrà sempre dire che quell'URL "http://www.miosito.mio/posto_libero" è equivalente a (o <sameAs>) "http://www.ontologiastandard.net/posto_libero".

...E ora?

L'Italia ha un grande bisogno di attivare fattori di crescita economica e sociale. La definizione di una solida e coraggiosa politica sulle informazioni del settore pubblico è forse, tra tutte, l'azione di governo in assoluto con il miglior rapporto costi-benefici, nonché quella che, tra tutte, dovrebbe registrare la più ampia convergenza tra tutte le parti sociali. Non di rado, attivisti open data e dipendenti pubblici volenterosi si offrono addirittura di iniziare progetti pilota su base volontaria: l'espressione di una volontà politica favorevole ai dati aperti è, spesso, tutto ciò che manca e l'investimento chiave da effettuare... per cui, **non resta che provarci!**

Il processo di apertura dei dati è ancora all'inizio nel nostro paese. L'unica – provvisoria – conclusione è quella di citare Tim Berners-Lee, chiedendo "DATI – GREZZI – ORA!" e con licenze libere.

Numerosi esempi ispiratori potranno aiutarvi a non partire da zero e a contare sugli approfondimenti di cui avete bisogno. A questo proposito, ecco un elenco di riferimenti utili.

Amministrazioni che fanno Open Data

- All'indirizzo http://www.dati.gov.it/content/ infografica è disponibile una ricognizione delle PA (dal Comune di Firenze alla Regione Emilia Romagna, passando per Piemonte, Lombardia e Veneto, per citarne solo alcune) e di altri organismi (come Istat e CNR) che fanno Open Data in Italia.
- All'indirizzo http://g.co/maps/xhz59 è disponibile una mappa aperta delle leggi regionali in materia di Open Data.

Comunità di attivisti Open Data

- http://www.spaghettiopendata.org/, nella cui mailing list (https://groups.google.com/forum/?fromgroups#!forum/spaghettiopendata) potrete trovare interlocutori con interesse (e passione) sul tema.
- Il ramo italiano della Open Knowledge Foundation (http://it.okfn.org/), anch'esso dotato di una nutrita mailing list (http://lists.okfn.org/mailman/listinfo/okfn-it).
- La sezione di InnovatoriPA dedicata ai dati aperti (http://www.innovatoripa.it/category/argomenti/open-data)
- http://www.datagov.it/, Associazione Italiana per l'Open Government.
- L'associazione Linked Open Data Italia, focalizata sulle tecnologie del web semantico, http://www.linkedopendata.it/
- L'associazione Wikitalia, http://www.wikitalia.it/

Manuali Open Data

L'Open Data Handbook della Open Knowledge Foundation, http://opendatahandbook.org/en/

Come si fa Open Data?, a cura della Associazione Italiana per l'Open Government, http://www.datagov.it/2011/05/11/come-si-fa-open-data-ver-2-0/

Qualche esempio di riutilizzo di dati pubblici

- Alcune applicazioni sviluppate a partire dai dati messi a disposizione in Regno Unito, http://data.gov.uk/apps
- Alcune applicazioni sviluppate a partire dai dati messi a disposizione in USA, https://explore.data.gov/catalog/apps/
- Altri esempi di riutilizzo raccolti da Epsiplatform, http://epsiplatform.eu/products_services
- L'Osservatorio ICT Piemonte ha realizzato un interessante studio che analizza i modelli di business basati sulla PSI. Il report integrale è disponibile al seguente indirizzo: http://bit.ly/HmGhks.

Progetti su temi PSI e Open Data

- Extracting Value from Public Sector Information, http://www.evpsi.org/
- Legal Aspects of Public Sector Information, http://www.lapsi-project.eu/
- Opening Data Architectures and Infrastructures of European Public Administrations, http://www.open-dai.eu/
- Ricognizione e analisi dei portali Open Data in Europa nell'ambito del progetto POPSIS, http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/11_2012/open_data_portals.pdf

Bibliografia

La bibliografia di riferimento di questo libro è consultabile all'indirizzo http://www.evpsi.org/biblio.

Questo libro è aperto ai vostri commenti fino al 15 giugno 2012!

Qualora intendeste suggerire modifiche, fornire nuovi riferimenti o, semplicemente, confrontarvi con noi in merito a specifiche questioni, non esitate a contattarci!

Federico Morando: federico.morando@gmail.com
Raimondo Iemma: raimondo.iemma@gmail.com
Claudio Artusio: claudioartusio@gmail.com

Gli autori ringraziano Michele Barbera, Alberto Cottica e Luigi Reggi per i preziosi commenti sinora forniti.